

# Bias in Voting Behaviour: Endogenous and Exogenous Factors

Zamira Noh

A thesis submitted in partial fulfillment of the requirements of the University of Lincoln  
for the degree of Doctor of Philosophy.

June 2017



## Abstract

Despite the vast research on the social bias in decision-making, relatively little is known about biases in voting behaviour. The main aim of this research was to explore alternative indirect methods to observe biases in decision-making and voting behaviour. A proximity bias was first observed in the rather unusual setting of the *Weakest Link* TV game show, when contestants avoided casting negative votes against their closest neighbours. This proximity bias was most profound for the contestant closest to the voter. Two field experiments were designed to test whether this *Neighbour Effect* occurred in different social contexts, among the first-year undergraduate students. The first study asked first-year undergraduate students in a lecture (n=449) to vote for another person seated in the same row. The same *Neighbour Effect* occurred when the vote carried a *nasty* (negative) outcome for the recipient however, when the vote valence changed to a *nice* (positive) outcome the *Neighbour Effect* disappeared. In negative voting, the result of the field experiment confirmed the original observation in the *Weakest Link*. However, a reverse polarity voting pattern was also found in the positive voting. This suggests participants significantly favoured their closest neighbour(s). The second field experiment used *Prisoner's Dilemma* with undergraduates in a lecture theatre (n= 229) to test the *Neighbour Effect*. The undergraduates played the game with another player seated in the same row and in the same block in a lecture theatre. The results showed a *neighbour effect* because the players were significantly more likely to cooperate with a neighbour than a non-neighbour. To conclude the findings from this study suggested that the *Neighbour Effect* is a robust bias in strategic decision-making and voting.

**KEYWORDS:** Decision-making, Neighbour Effect, Vote Valence, Voting, Prisoner's Dilemma

## **Dedication**

Throughout my life, there was one person who has always been there to support and listened during the difficult times without passing any judgement. This thesis is dedicated to my loving late grandmother, Hamsiah Ahim who always be in my heart. Although our time together was brief, her contribution to my life will be felt forever.

To my parent, thank you...

# Acknowledgements

“I think about you, night and day. I see you in my dreams. I sing your praises far and wide.

I write odes to you, no one will ever read. I weep at your absence.

I count the years until you’re mine” (The PhDComics, 2011)

Alhamdulillah, praise to Allah I have finally completed my thesis, with positive assumption that I will, in fact, complete this thesis. As of now, whether this enthusiasm will bear fruit and prove to be validated, no one can say beyond any doubt.

Firstly, I would like to express my sincere gratitude to my advisor, Dr. Paul Andrew Goddard for the continuous support for my PhD study and related research, for his patience, motivation, understanding and immense knowledge. His guidance helped me in all the time of the research and writing of this thesis. I seriously could not have imagined having a better advisor and mentor for my PhD study.

Besides my advisor, I would like to thank my transfer viva committee: Dr. Fenja Ziegler and Dr. Petra Pollux for their insightful comments and encouragement, but also for the hard questions (yes, I cried!) which incited me to widen my research from various perspectives.

My sincere thanks also goes to the Identity and Social Relationships research group members, Dr. Rachel Bromnick, Dr. Patrick Hylton, Dr. Adrian Parke, Dr. Ava Horowitz, Dr. Roger Bretherton, Dr. Ruth Gaunt, Professor Harriet Gross, Dr. David Ellis and

Mariana Lobo Pinho who provided me an opportunity to join their team as a research group member. I enjoy every bit of the meetings and luncheons!

Special thanks to the School and College administration, Zoe Mead, Ian Brown and Nicola Knaggs for their precious support in easing the documentation processes, the technician team, Christopher John Luke, Heather Shaw and Ferenc Igali who constantly helped with my weird enquiries and of course, giving the access to research facilities and equipment. And, to my subject librarian, Oonagh Monaghan who always be there to help me with any unsolved issues with the library. Without their precious support, it would not be possible for me to complete this thesis.

Also, I thank my friends who I met during my PhD study at the University of Lincoln and at the conferences for the stimulating discussions, the support, laughter, patience, encouragement, meals, coffees, hugs and fun we have had in the last three years.

Last but not the least, I would like to thank my family: my parent, my brothers, sister, grandpa, uncles and aunts for supporting me spiritually throughout writing this thesis and my life in general.

# Table of Contents

Abstract	iii
Dedication	iv
Acknowledgement	v
List of Abbreviation	x
List of Tables	xi
List of Figures	xii
List of Appendices	xvi
Thesis Organization	1
 <b>Chapter 1 Introduction</b>	 <b>5</b>
<b>I1 Method used to study “Nice” and “Nasty” Behaviour</b>	<b>6</b>
I1.1 Chapter Outline	6
I1.2 Methods in Social Psychology	7
I1.3 Why We do Nasty Things to Other People? A Review of Milgram (1963) Experiment.	13
I1.4 The Prisoner’s Dilemma	20
I1.5 Chapter Summary	25
 <b>I2 Evaluation of The Weakest Link TV Game Show</b>	 <b>30</b>
I2.1 Chapter Outline	30
I2.2 Using the Weakest Link Game Show to Test for Discrimination, Strategy and Proximity.	31
I2.3 Evaluation of the Weakest Link Game Show Studies	54
I2.4 Research Aim and Objectives	58

<b>Chapter 2</b>	<b>Methods</b>	<b>59</b>
	<b>M1 General Methods</b>	<b>60</b>
	M1.1 Chapter Outline	60
	M1.2 Observational Field Study on the Weakest Link TV Game Show	61
	M1.3 Pilot Studies	66
	M1.4 Lecture based Field Experiments	69
<b>Chapter 3</b>	<b>Results</b>	<b>78</b>
	<b>R1 Demonstration and Replication of The Neighbour Effect in The Weakest Link TV Game Show using Observational Field Studies</b>	<b>79</b>
	R1.1 Chapter Outline	79
	R1.2 Observational Field Study 1: Original Demonstration of the Neighbour Effect	81
	R1.3 Observational Field Study 2: Replication of The Neighbour Effect	116
	R1.4 Observational Field Study 3: Analysis of Second Round Voting Reciprocity, Retaliation and Consistency	132
	R1.5 General Discussion	150
	<b>R2 Testing the Neighbour Effect using Voting Activity</b>	<b>156</b>
	R2.1 Chapter Outline	156
	R2.2 Introduction	158
	R2.3 Method	163
	R2.4 Results	169
	R2.5 Discussion	184



<b>R3</b>	<b>Testing the Neighbour Effect using Prisoner's Dilemma</b>	<b>187</b>
R3.1	Chapter Outline	187
R3.2	Introduction	189
R3.3	Method	196
R3.4	Results	204
R3.5	Discussion	211
<b>Chapter 4</b>	<b>Discussion</b>	<b>213</b>
<b>D1</b>	<b>General Discussion</b>	<b>214</b>
D1.1	Chapter Outline	214
D1.2	Summary of the Results	215
D1.3	Evaluation of Main Contributions	220
D1.4	Limitations of the Study	229
D1.5	Suggestions for Future Study	232
D1.6	Conclusion	234
	<b>Appendices</b>	<b>235</b>
	<b>References</b>	<b>259</b>

## List of Abbreviations

<b>CG</b>	Coins Game
<b>CLT</b>	Construal Level Theory
<b>CP</b>	Credit Point
<b>GBS</b>	Group Bias Statistic
<b>NE</b>	Neighbour Effect
<b>PD</b>	Prisoner's Dilemma
<b>WL</b>	Weakest Link
<b>15-to-1</b>	Fifteen to One

## List of Tables

### Chapter 2

Table 1	The Money Tree for the Daytime Episodes	62
Table 2	The Type of Votes Performed by the Voter	71

### Chapter 3

Table 3	The Calculated Expected Frequencies for Each Voter-Votee Position	87
Table 4	Contestant Arrangement by Gender (72 - episode Study)	95
Table 5	The Contestants Consensus Votes in 72 Episodes	99
Table 6	The Calculated Frequencies for Null Consensus for Different Level of Consensus	104
Table 7	The Calculated Frequencies for Voting by Consensus (72 – episode Study)	105
Table 8	The Calculated Expected Frequencies for 151 Episodes	119
Table 9	The Chi-Square Value for Each Spatial Position.	121
Table 10	Contestant Arrangement by Gender (151- episode study)	126
Table 11	The Contestants Consensus Votes in 151 Episodes	127
Table 12	The Contestants Consensus Votes in 295 Episodes	147
Table 13	The Participants' Gender (n=449) for 2013/2014 and 2015/2016 Cohorts	163
Table 14	The Observed Seats Available for Each Row by Block in the Lecture Theatre	166
Table 15	The Expected Frequencies for the Voting Activity by Cohort	168
Table 16	The Overall $\chi^2$ Value for Each Spatial Position for 2013/2014 and 2015/2016 Cohorts	171
Table 17	The $\chi^2$ Value for Each Spatial Position over Voting Valences for 2013/2014 and 2015/2016 Cohorts	175
Table 18	The Participants' Demographics (n=229) for Cohort 2014/2015	197
Table 19	Original Table Score Presented in the Response Sheet	198
Table 20	The Distributions of the Coin Game Response Sheet	199

# List of Figures

Figure 1	The thesis organization by chapters	4
<b>Chapter 1</b>		
Figure 2	The original Prisoner's Dilemma payoff table presented for two-person game	20
<b>Chapter 2</b>		
Figure 3	The other contestants' position based on contestant A as the <i>voter</i>	63
Figure 4	The observation protocol for the Weakest Link TV game-show	64
Figure 5	The lecture theatre seating arrangement in the University of Lincoln	70
Figure 6.1	The instruction given to the participants on how to take part in the voting task for 2013/2014 cohort	73
Figure 6.2	The instruction given to the participants on how to take part in the voting task for 2015/2016 cohort	73
Figure 7	The payoff matrix for the coin's game	74
Figure 8	The instruction given to the participants on how to take part in the coin's game for 2014/2015 cohort	75
<b>Chapter 3</b>		
Figure 9	Weakest link as a function of contestant position.	88
Figure 10.1	Voting frequency as a function of spatial position for 72-episode study (including the votes cast by eventual weakest link)	90
Figure 10.2	Voting frequency as a function of spatial position for 72-episode study (excluding the votes cast by eventual weakest link)	90
Figure 11	The vote frequencies with respect to female and male contestants for 72-episode study.	92
Figure 12.1	Voting performance for female contestants in 72-episode study	93
Figure 12.2	Voting performance for male contestants in 72-episode study	93
Figure 13.1	Votes by gender as a function of spatial position for female voters in 72-episode study	98
Figure 13.2	Votes by gender as a function of spatial position for male voters in 72-episode study	98

Figure 14	The comparison between the observed 72-episode and the null hypothesis distribution of votes	106
Figure 15	The vote ratio as a function of proximity (72-episode study)	107
Figure 16	The vote ratio as a function of proximity by consensus	108
Figure 17.1	Vote frequency as a function of spatial position (consensus 6,7, and 8)	110
Figure 17.2	Vote frequency as a function of spatial position (consensus 2,3,4 and 5)	110
Figure 18	Vote frequency as a function of spatial position with error bars	119
Figure 19	Weakest link as a function of contestant position for the first round (151-episode study)	120
Figure 20.1	Voting frequency as a function of spatial position for 151-episode study (including the votes cast by eventual weakest link)	122
Figure 20.2	Voting frequency as a function of spatial position for 151-episode study (excluding the votes cast by eventual weakest link)	122
Figure 21	The vote frequencies with respect to female and male contestants for 151-episode study.	123
Figure 22.1	Voting frequency as a function of spatial position for 151-episode study (including the votes cast by eventual weakest link)	125
Figure 22.2	Voting frequency as a function of spatial position for 151-episode study (excluding the votes cast by eventual weakest link)	125
Figure 23	The comparison between the observed 151-episode and the null hypothesis distribution of votes	128
Figure 24	The vote ratio as a function of proximity (151-episode study)	129
Figure 25	Weakest link as a function of contestant position for the second-round	136
Figure 26.1	Voting frequency as a function of spatial position from the second-round voting (72-episode study)	137
Figure 26.2	Voting frequency as a function of spatial position from the accumulated votes for first and second round of 72 episodes and 151 episodes	137
Figure 27	Breakdown of contestants in Venn diagram in four categories	140
Figure 28.1	Second round voting of the potential to retaliate contestants	142

Figure 28.2	Second round voting of the potential to be consistent.	142
Figure 29.1	Second round voting of the potential to retaliate versus non-retaliation	145
Figure 29.2	Second round voting of the potential to be consistent versus non-consistent	145
Figure 30	The comparison between the observed 295-episode and the null hypothesis distribution of votes	147
Figure 31	The vote ratio as a function of proximity (72-episode and 151-episode studies)	148
Figure 32	Overall vote frequency as a function of spatial proximity; proximal, medial and distal	172
Figure 33	Overall vote frequency as a function of spatial proximity for 2013/2014 and 2015/2016 cohorts	173
Figure 34	Overall negative voting from 2013/2014 and 2015/2016 cohorts	176
Figure 35	Overall negative voting by different cohorts, 2013/2014 and 2015/2016	177
Figure 36	Overall positive voting from 2013/2014 and 2015/2016 cohorts	178
Figure 37	Overall positive voting by different cohorts, 2013/2014 and 2015/2016	179
Figure 38	Neutral voting from 2013/2014 cohort	180
Figure 39	Positive voting at +1 and +5 valences in 2013/2014 cohort	182
Figure 40	Negative voting at +1 and +5 valences in 2013/2014 cohort	183
Figure 41	The gain-frame and loss-frame pay-off structures for the Prisoner's dilemma and the Chicken game	192
Figure 42	The payoff matrix for the coin's game	195
Figure 43	Diagram of version 1 and version 1r with the seat arrangement in the lecture theatre	200
Figure 44	Diagram of version 2 and version 2r with the seat arrangement in the lecture theatre	201
Figure 45	Diagram of version 3 and version 3r with the seat arrangement in the lecture theatre	202
Figure 46	Illustration of player <i>Y</i> pathway to play against player <i>X</i>	203
Figure 47	Participant's knowledge about the activity	205

Figure 48	Participants' response against their opponent's choices	206
Figure 49	Response rate as a function of recipients' position	207
Figure 50	Vote frequency as a function of spatial proximity when player $X$ is at lateral, longitudinal and both (lateral and longitudinal) positions	209
Figure 51.1	Vote frequency for the direct neighbour(s) as a function of spatial proximity	210
Figure 51.2	Vote frequency for the indirect neighbour(s) as a function of spatial proximity	210

## List of Appendices

A	The generic background and instructions for 2013/2014 cohort	236
A1	First page for negative vote and removes one voucher	237
A2	Second page for negative vote and removes one voucher	238
A3	First page for negative vote and removes five vouchers	239
A4	Second page for negative vote and removes five vouchers	240
A5	First page for positive vote and adds one voucher	241
A6	Second page for positive vote and adds one voucher	242
A7	First page for positive vote and adds five vouchers	243
A8	Second page for positive vote and adds five vouchers	244
A9	First page for neutral vote and gives zero voucher	245
A10	Second page for neutral vote and gives zero voucher	246
B	The generic background and instructions for 2015/2016 cohort	247
B1	First page for negative vote and removes five vouchers	248
B2	Second page for negative vote and removes five vouchers	249
B3	First page for positive vote and adds five vouchers	250
B4	Second page for positive vote and adds five vouchers	251
C	The generic background and instructions for 2014/2015 cohort	252
C1	Version 1 of the Coins game activity	253
C2	Version 1r of the Coins game activity	254
C3	Version 2 of the Coins game activity	255
C4	Version 2r of the Coins game activity	256
C5	Version 3 of the Coins game activity	257
C6	Version 3r of the Coins game activity	258



## Thesis Organization

The overall layout for the thesis is shown in Figure 1 below. It illustrates the thesis organization, in four main parts that follow the typical structure for an empirically based scientific report: **Introduction**, **Methods**, **Results** and **Discussion**.

The subject matter of the thesis utilises a field of enquiry that has previously received relatively little attention in the past in the field of Psychology. The **Introduction** therefore presents a general background and literature review focusing on an evaluation of different methods used to assess social interactions from different disciplines to understand social behaviour. The second chapter in the **Introduction** provides a more detailed evaluation of studies that have made use of the Weakest Link (WL) TV game-show. It presents the details of the observational field studies used to assess the voting behaviour of the contestants in the UK version of the TV game-show the WL.

The **Methods** also detail the experimental work used to test the findings from the observational field studies. The **Results** are in two main parts detailing the findings using the two main methods. The first part presents the findings from the observational field studies in three sections. Similarly, the second part of the **Results** shows the findings of the experimental work in two sections. Finally, the **Discussion** considers the results in the context of relevant previously published materials. The contents of the chapters are presented in more detail below.

**Introduction Chapter** deals with the problem of limitation of understanding social behaviour and shows how the alternative approach has been developed and presented in two parts. The first part (**I1**) presents the dilemma in experimental Social Psychology and critique of the infamous classic social experiment by Milgram (1963). It reviews approaches used in two different disciplines where the similarities and the differences in methodological approaches were clarified. In Social Psychology, it focuses on experimental procedures, however in Economics, it took a stand to understand the behaviour in a natural setting as in a form of game theory. Hence, this chapter elaborates on how the two distinctive disciplines can explain social behaviour. The second part (**I2**) presents a critical evaluation of studies that have employed observations of the WL game-show. Finally, the rationale behind the studies, together with aims and objectives of this thesis are presented.

**Methods Chapter** describes the methods used in this thesis. The first method (**M1.2**) focused on an unorthodox approach in Social Psychology, which is an observation of a TV game-show. The second method (**M1.4**), emphasises the field experiment studies used to test the findings found from the initial study. The pilot studies (**M1.3**) were briefly discussed.

**Results Chapter** presents the results in the form of three studies. **R1** presented the demonstration and replication of the original observation of the contestants' voting behaviour on the WL. **R2 and R3** are presented two main large scale lecture studies that were used to test the findings from the observations.

**Discussion Chapter** discusses the findings in general in response to the aim of the research. It gives the implications which are considered to be significant and relevant for academia and society in general.

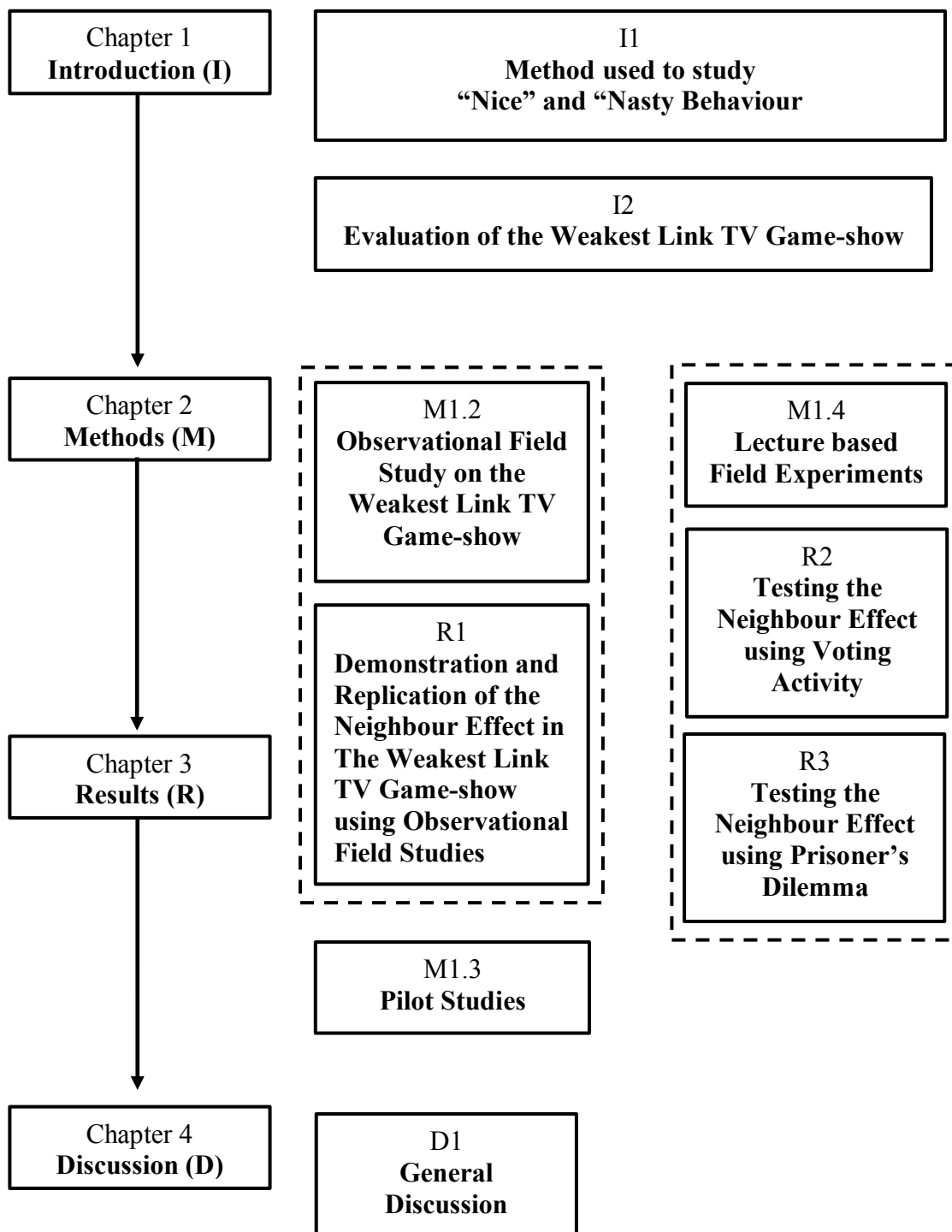


Figure 1. The thesis organization by chapters.

# Chapter 1

## Introduction

The starting point for this research was a simple observation based on a popular TV game-show, The Weakest Link (WL). Contestants on the show were placed in a unique social dilemma by being required to name one of their fellow contestants as the worst performer, the so-called “weakest link”. This thesis is concerned with the endogenous and exogenous factors in voting behaviour.

The **Introduction** is presented in two main parts. The first part, **I1** evaluates the general methods in Social Psychology and puts the WL game-show in the context of academic research making the case for considering the format of TV game-shows as an important resource methodology for testing hypotheses in Psychology, Economics and Social Science. The argument is made that the decisions made by the contestants on the TV game-show can be operationally defined as a basis to study how people decide to do something “bad” or “nasty” to another person. This allusion is considered within the context of Milgram’s (1963) famous study on obedience and the Prisoner’s Dilemma game.

The second part, **I2** considers how the WL game-show has been used to test theories of discrimination, strategic game playing and proximity bias in decision-making. The chapter finishes with a statement of the research goals.

## **Chapter I1:**

### **Methods used to study “Nice” and “Nasty” Behaviour**

“One of the worst thing in life is not how nasty the nasty people are. You know that already. It is the how nasty the nice people can be” (Powell, 1962, p. 246).

#### **I1.1 Chapter Outline**

This first chapter makes the case that the voting behaviour of contestants on the WL TV game-show can be used as a way to study of how people choose whom to perform a “nasty” or “bad” act on.

The chapter is structured by reviewing the general methods used in Social Psychology and puts WL in the context of an alternative research methodology. Milgram’s (1963) classic work on obedience is reviewed as is the Game Theoretic approach using the example of the Prisoner’s Dilemma (PD). Comparisons are made between Milgram’s work, PD and WL using strategic features such as “nice” and “nasty” strategies, cooperation and conflict/defection.

I1.1 Chapter Outline

I1.2 Methods in Social Psychology

I1.3 Why We do Nasty Things to Other People: A Review of Milgram (1963) Experiment.

I1.4 The Prisoner’s Dilemma

I1.5 Chapter Summary

## **I1.2 Methods in Social Psychology**

**The dilemma of Social Psychology.** The social psychologist seeking to examine how people make social choices faces a conundrum similar to that of the cognitive psychologist endeavouring to understand human consciousness. Both psychologists spend their academic and personal lives seemingly completely immersed by their subject matter, yet once subjected to scientific scrutiny the essence of their specialism escapes simple description remaining tantalizing and elusive. One reason for this is that despite the fact we are completely enveloped by our social milieu and conscious experience, the milieus and conscious states are dynamic and multifaceted. Another reason is that the individual conscious social actor is not always rational or fully aware of the factors that affect their social experience. Humans are notoriously prone to a wide range of biases both in their decision-making and their social judgments regarding others and themselves (Kahneman & Tversky, 1972; Tversky & Kahneman, 1975). Furthermore, take the myriad decisions people routinely make that can range in importance from the life changing and profound, choosing whether or not to have a child with a partner, through to the trivial, for example, what to watch on TV tonight.

Given the complexity of the subject matter it is not surprising that several methods have been adopted to study human social behaviour and social cognition which can be subdivided broadly into the more quantitative experimental social psychological approach based on logical positivism and the more qualitative constructive approach founded on discourse and narrative accounts (Bryman, 1984). The former is open to criticism because the behaviour of participants knowingly taking part in an experiment is not the same as

their social decisions made in everyday life. Social behaviour by its very nature is sensitive to the context in which it occurs and so the psychology laboratory becomes a new kind of social environment in its own right. By directly *looking* at social behaviour, the social psychologist faces the charge of changing that behaviour. Equally, qualitative methods can be used to generate richer accounts of social interactions yet face the criticism of lacking the methodological rigor and control lauded by the experimentalist. Adopting either research strategy, of course, affords benefits and insights but these come at some cost. Taking a nomothetic approach to identify axiomatic laws for human social behaviour loses the experience of the individual agent while the converse is also true because an idiographic approach can provide a detailed understanding of an individual agent but one which may not be applicable to the population as a whole.

The trade-off between using nomothetic and idiographic methods to understand how an individual makes a choice was cleverly captured by Sherlock Holmes in ‘The Sign of The Four’:

“...He remarks that, while the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty. You can, for example, never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So, says the statistician” (Doyle, 1890, p. 169).



**An alternative methodology: How watching TV game-shows can tell us about human nature?** An alternative methodology to those alluded to above is used in the first part of this thesis, a field observation of contestants on a TV game-show. TV game-shows have long been used for material by Economics researchers, although their use in Psychology has been more sparing.

Metrick (1995) argued that TV game-shows provide a suitable empirical resource to analyse behaviour, as many of the shows are structured as well-defined decision problems. The TV game-show in question is the WL (i.e., see **Methods** section **M1.2**, p. 60) for a detailed breakdown of the rules of the game). This game-show involves several interesting topics such as voting consideration, discrimination, strategy optimization and the existence of Nash equilibria. The successive elimination of the “weakest link” by vote became the distinctive feature of the show and gave rise to its distinctive name. It is of interest because each of the nine contestants that took part in an episode (i.e., in the UK version) faced a dilemma because they had to make a fixed judgment about one of their fellow contestants. It became relatively straight forward thereafter to observe and measure their social judgments in groups that would otherwise not be observable in more naturalistic settings and that would be almost impossible to replicate in the laboratory. The feature of the show that was of particular relevance for the work in this thesis occurred when the contestants had to pick a fellow contestant to be eliminated from the show as the worst performer in the group following a round of quiz questions.

*Nasty voting.* The judgment required from the participants was also particularly interesting, because, following a first three-minute round of general knowledge questions, they each had to nominate, openly, which of their fellow contestants was the worst performer, the so-called weakest link. It is fairly uncommon to be able to find and measure instances of disapproval voting such as this. Another way of interpreting the decision each contestant had to make is that it was a bad or nasty judgment to bestow on another contestant. Ordinarily in social situations people are more likely to follow the conventions, manners and social norms of their culture by being predisposed to be openly pro-social. Here contestants publicly denounced one of their number as the weakest link. For this thesis, the vote each contestant made was operationalized as an action by doing a “negative”, “bad” or “nasty” thing to another person. The tag of nasty is used as a way of contrasting it with the nice strategies referred to later in the thesis with respect to game theory. Ball (2004), noted that one set of strategies used in the PD (see later for fuller discussion) were considered nice if they led to cooperation with an opponent and suggested that the alternative strategies could be tagged as nasty.

Another reason for considering the vote as nasty is because the contestant that subsequently received most nominations was eliminated from the game. The remaining eight contestants then proceeded to another round of general knowledge questions. The assumption is made that a rational player would avoid elimination and so recognize that casting a vote for another player is nasty because it increases their probability of being eliminated from the game.

*Methodological advantages of using WL.* Analysing the voting patterns of contestants on the WL game-show therefore provided a means to study how people made nasty decisions. The format of the TV game-show also conferred several advantages methodologically.

1. Participants, or contestants, believe they are playing a game on TV game-show rather than being studied regarding their decision-making in a psychological laboratory. Their behaviour is likely to be more natural than if it occurred in a laboratory.
2. Participant/experimenter effects. Both participants and experimenters are prone to bias in experimental situations. Such biases can be either or both conscious or unconscious. Participants are vulnerable to using implicit social cognition to interpret the aims of an experiment and seek to conform to, or contradict, the perceived aims. Experimenters, similarly, are prone to bias to confirm hypotheses. As participants are contestants on a TV game-show they are naive to the aims and purpose of the research. Experimenters were removed from the design, participant recruitment and execution of the TV game-show so had no influence on the procedure and were unable to introduce unintended systematic bias into the procedure.
3. Randomisation. Double-blind procedure. The research protocol is effectively double-blind as participants and researchers were naive to the research aims and participant allocation respectively. Independent observers were used to code data.

4. Clearly defined variables. The dependent variable in the TV game-show was the vote choice made by each contestant. Independent variables included the relative positions of the candidates and their demographic background (e.g., gender).
5. Ethics. Studying the responses made by contestants on TV game-shows are free from the kinds of ethical concerns that would make their replication difficult to achieve in the psychology laboratory. The threshold for what TV production companies would consider acceptable is very different to the code of ethics set out by the British Psychological Society (Broadcasting Act 1996).

### 11.3 Why We do Nasty Things to Other People? A Review of Milgram (1963) Experiment.

“... (it is) a social fact; the individual who is commanded by a legitimate authority ordinarily obeys. Obedience comes easily and often. It is a ubiquitous and indispensable feature of social life” (Milgram, 1963, p. 372).

A series of now classic, seminal social psychology experiments still resonate in modern psychology largely because the behaviour of the participants was so unexpected. One of the most controversial demonstrated that ordinary participants under laboratory conditions could be cajoled into delivering what they believed to be extreme and lethal electric shocks to another person (Milgram, 1963). Participants were led to believe that they were volunteering to take part in a study investigating the effect of punishment on learning. They were allocated the role of a *teacher* in a mock learning task and administered a “punishment” by delivering electric shocks to an errant *learner*. The learning task, as it turned out, was a sham, a carefully orchestrated hoax designed to be plausible to the unsuspecting participant. The *learner*, of course, was Milgram’s stooge, an actor following a pre-set script, fully briefed and aware of the proceedings and aware of the fact that no electric shocks were given. The real purpose of the study was to measure the maximum electric shock that the participant was prepared to go to. Shock magnitude began at 15v and increased by increments of 15v up to a maximum of 450v. The participants bought into the conceit and believed that they were administering real electric shocks to a real *learner*. Milgram reported that the participants frequently displayed extreme agitation in the face of having to inflict pain and harm to the *learner*. Nevertheless, in the original study, 24 of the 40 participants went up to the maximum

450v. The remaining 16 terminated the study by refusing to go to the maximum, but they still went up to at least 300v, sufficient to be lethal or at least cause serious discomfort. The experiment has been repeated and replicated using different groups (i.e., up to 780 participants) Milgram (1963, 1964, 1965). The participants reported that they were under extreme emotional stress after the experiment, even though majority of the participants also reported that they appreciated being part of the experiment, "...84% of the subjects stated they were glad to have been in the experiment..." (Milgram, 1964, p. 849). Even though Milgram's study received widespread criticism on the ethical consideration towards his participants (Baumrind, 1964), the findings still resonate throughout Psychology, not least because it possibly revealed an inconvenient truth about human nature.

*The impact of Milgram's study.* There are many ramifications of Milgram's (1963) famous study but it remains the case that seemingly normal people could be placed into a situation that allowed them to do something nasty to another person. In this case, the nasty behaviour was delivering a very nasty electric shock. Moreover, Milgram was able to entice this behaviour out of the participants in a very short period of time, usually less than 50 minutes. If nasty behaviour can be so readily elicited in normal, nice people, by Milgram under the subterfuge of a psychology experiment, then those powers intent on engineering systematic harm and genocide on entire populations could similarly recruit obedient agents to carry out their bidding.

Milgram (1963) made reference to the need for obedience to enable the mass killings in death camps in Europe that occurred on an industrial scale around the time of the Second World War. Similarly, Arendt's (1964) record of Eichmann's trial in

Jerusalem, detailed how Eichmann constructed, maintained and managed a ruthlessly efficient system to enable the mass slaughter that was perpetrated. This smooth operation required the complicity of hundreds of thousands of people to make it work. Eichmann, rejected the charge of murder levelled against him maintaining that he had never killed a human being in his life. If anything, he should only be accused of “aiding and abetting” (Arendt, p. 22) as the charges of murder made against him were *acts of state* and not crimes that can be attributed to an individual. According to his lawyer “it had been his duty to obey” (Arendt, p. 22 Arendt). Arendt subtitled her work ‘A report on the banality of evil’ to reflect the calm, detached and routine fashion by which Eichmann engineered the mass killings. Milgram’s study can be used as a demonstration that people can be cajoled easily to carry out “nasty” acts.

More recent texts suggest that Arendt’s account missed out on Eichmann’s true motives (Stangneth, 2014). Eichmann simply attempted to promote the idea that he did not fit the profile of a mass murderer but did fit the profile of an obedient and efficient administrator. Recently reviewed documents of his time in Argentina, prior to his trial in Jerusalem, show a more calculating architect of the holocaust fully endorsing its goals.

Milgram’s works and its implications not only caused surprise in the academic world, but provoked the development of a code of ethics for the treatment of human participants in psychological research. It raised concerns about the conflicting interests between researchers and participants. “The *game* is defined by the experimenter and he makes the rules” (Baumrind, 1964, p. 421). Baumrind stated that Milgram created unacceptable levels of stress towards his participants. Furthermore, Perry (2012)

suggested that Milgram's participants were not properly debriefed, something Milgram persistently refused to acknowledge.

The recorded interviews, highlighted by Perry (2012), which followed the experiment suggested that the participants were frequently left with conflicted feelings and had expressed their discomfort about their participation. However, conversely, Brannigan (2013) suggested that the work carried out by Perry was not sufficient to make out that any of the former participants had indeed been mistreated, so this critique of Milgram's experiment was unwarranted. Nevertheless, partly as a result of the impact Milgram's study had on its participants, studies thereafter have had to adhere to explicit ethical guidelines as laid down by the relevant professional bodies (e.g. American Psychological Association and British Psychological Society).

***Critique of Milgram's controversial experiment.*** Despite its critics, the experiment inspired many researchers to replicate the methodology in a variety of experiments (Burger, 2009; Kilham & Mann, 1974; Meeus & Raaijmakers, 1995; Sheridan & King, 1972; Slater, Antley, Davison, Swapp, Guger, Barker, . . . Sanchez-Vives, 2006). The infamous experiment was also recognized world-wide in the form of documentaries such as *Curiosity: How Evil are you?* by Eyres, Bowie, Smithson, Winslow and Day (2011), *The Heist* by Mills (2006) and *The Milgram Re-enactment* by Dickinson (2002). It has also been recognized in movies such as *Compliance* by Zobel, Lin, Sena, Muskat and Davidson (2012), *The Tenth Level* by Bellak (1976), *Experimenter* by Abeckaser, Golombek, Melita, Robbins, Schoof and Singer (2015) and a game-show (e.g., *Le Jeu de la Mort*, 2010). Milgram's study was not only relevant to understand the notorious conflicts



that surfaced during the post-world wars such as; Cold war, Vietnam war, Korean war, Gulf war, Iraq war or Cuban revolution, it was an emerging research area particularly to understand the extent to which people would go.

*Validity and Reliability.* Milgram (1963, 1965, 1974) suggested the existence of an *agentic state* in which the participants were made to believe that they were not responsible for any action carried out throughout the experiments. They believed that they were acting as an instrument to penalize the *learner* and acted the role accordingly, since the responsibility lay with the experimenter themselves. This disturbing finding showed a nasty part of the human unconscious state of mind where they deliberately agreed to administer harm, despite going against their conscious, rational conscience.

Brown (1986) stated that the disagreements among researchers revealed the opposite assumption of how a rational individual should behave. Brown therefore regarded the unexpected findings from Milgram's experiment as the most important psychological research. Also, Blass (1991, 1996, 1999, 2002, 2004, 2009) based a body of work on Milgram's experiment and stated that "...it has become one of the best-known works of social psychology in the world, as evidenced, for example, by the fact that Milgram's (1974) book has been translated into 11 different languages" (Blass & Schmitt, 2001, p. 115).

Milgram also argued that the *situation* was the more important determinant of social behaviour than the individual's personality. He repeated the experiments with various other conditions and found the level of obedience changed with the level of

physical and emotional attachment. When a condition where the *teacher* and *learner* were related, either romantically or by family, obedience dropped to the lowest obedience recorded (defiance was 85%) (Perry, 2012). However, Baumrind (1964) contested that the context of the laboratory setting was ambiguous for the participants. It was such an unfamiliar environment for the participants that they became even more compliant with authority figures by obeying, because of their uncertainty about how to behave in that situation. Blass (1991, 1996, 1999, 2002, 2004, 2009) also suggested that the key factor in Milgram's paradigm was the odd situation that the participant was suddenly placed in, such that strong characters in other social domains became uncertain and indecisive in the experiment.

Observations from experimental game literature also show that the status of the individual "punisher" is an important factor in how participants judge the success and appropriateness of a punishment meted out. For example, when participants were asked to judge the outcome of vignettes where social transgressors (Gordon, Madden & Lea, 2014; Gordon & Lea, 2016) were asked to rectify their behaviour by a third-party-punisher, they judged the likelihood of the transgressor remedying the transgression according to the perceived status of the transgressor and the third-party-punisher. High status third-party-punishers were predicted to cause the social transgressor to change their behaviour more than the lower status third-party-punishers. This could be interpreted in terms of the cost/benefit analysis of status and punishment from an evolutionary perspective.

Aside than the *status* of the individual "punisher", from the experimental studies, it showed that, the success of the "intervention" (i.e., to punish the aggressor) by the

dominant “punisher” (*high-status*) was not affected by the heterogeneity of the punishment or threat, either as aggressive (physical violence) or ostracism (non-aggressive) ways. As such, it suggested that it affects the “cost” of the punishment in which, the dominant “punisher” could ride on the benefit of the reputational gains, where the attempt of retaliation was lesser. Hence, these experimental studies concur with Milgram’s claims that the *status* of the individuals in the obedience to authority scenario was one of the factors that could regulate the effect of retaliation or cooperation.

In the following section the Prisoner’s dilemma game is described as a useful method to explore strategic behaviour.

## I1.4 The Prisoner's Dilemma

This is a standard method to test a situation when two players each have two options where the outcome depends on the choices made by the other. Prisoner's Dilemma (PD) is one of the best-known game strategies in social science.

		II	
		confess	not confess
I	confess	$(-1, -1) \textbf{PP}$	$(1, -2) \textbf{TS}$
	Not confess	$(-2, 1) \textbf{ST}$	$(0, 0) \textbf{RR}$

Figure 2. The original Prisoner's Dilemma payoff table presented for two-person game (non zero-sum game). The pay-off matrix is reprinted from "The Mathematics of Tucker: A Sampler" A. W. Tucker, 1983, *The Two-Year College Mathematics Journal*, 14 (3), 228. Copyright (1983) by Mathematical Association of America. **P** is punishment **T** is temptation, **S** is sucker and **R** is reward.

It was originally framed by Merrill Flood and Melvin Dresher in 1950 as part of RAND Corporation project to investigate a game theory, where it was later formalised by Albert William Tucker in 1983 (Poundstone, 1992), in which he expressed the situations in a simple pay-off table. As Figure 2 shows a simplified dilemma faced by two prisoners in which they were given options either to confess or not confess, where it is also considered in terms of cooperate and defect.

In this scenario, the prisoners were told;

- “1. If one confesses and the other does not, the former will be given a reward of one unit and the latter will be fined two units,
2. If both confess, each will be fined one unit,
3. If neither confesses, both will go clear” (Tucker, 1983, p. 228).

Given these options, the prisoners had to decide whether to cooperate with their fellow prisoner, by not confessing, or defect with the other prisoner by confessing. However, neither of them knows what the other decided, which brings them to face with the dilemma of whether it is best to cooperate or defect.

Breaking the PD down it can be seen that the choices that the prisoners make will have a profound impact on the relative penalties that they could receive. For instance, if both prisoners, **I** and **II**, confessed (defected), both would be punished (**P**) by being fined one unit (NW corner in Figure 2). Whereas, if both cooperated, by not confessing, they would be rewarded (**R**) by escaping any fine (SE corner). The other possibility occurred when they differed in their choice to cooperate and defect. When one of the prisoners was the ‘sucker’ (**S**) by cooperating whilst the other succumbed to ‘temptation’ by defecting (**T**), the co-operator suffered the greatest penalty by being fined two units but the defector gained one unit (SW & NE corners). The PD is defined by the respective pay-offs such that  $T > R > P > S$ .

From a strategic point of view, it was always in the best interests of each player to defect by confessing. This is evident because whichever strategy the other prisoner plays the best outcome comes from defecting. Consider prisoner **I**, when **II** cooperated, **I** received one unit from defecting (confessing) but only got zero from cooperating (not confessing). When **II** defected by confessing, **I**'s best strategy was still to defect by confessing giving a penalty of -1. Cooperating led to a greater penalty of -2. Therefore, PD elegantly shows that when each individual pursued their own self-interest (defecting), the outcome is worse than when both cooperated. Technically, the rational *player* should always pick defection in PD because the rational individual seeks to maximise their utility. The unique features of PD, and its variants, that makes it so interesting from a research perspective is that the game has a unique solution by each player playing defect, whatever the other player does. However, the Pareto optimal solution for the game is for both players to play the opposite strategy as it was impossible to play without making the other player worse off, while the other better off. PD presents the dilemma in another sense as whether to play for one's self interest or whether to go for the overall best solution for all concerned. The dominant strategy of defection will be preferred to cooperation. However, the catch is, if both decided to cooperate by not confessing, it yielded the best outcome for both.

Notice that PD is a non-zero sum, non-cooperative game. It can be contrasted with pure competition zero sum games where there is always one winner and one loser. The pure strategy unique solution to the PD is referred to as the Nash equilibrium. It is the sole solution to the game. The Nash equilibrium can be defined as, "...an action profile  $a$  with the property that no player  $i$  can do better by choosing an action different from  $a_i$ , given

that every other player  $j$  adheres to  $a_j$ ". (Osborne, 2009, p. 22). Stated informally, this means that in the case of the example above with the two-player PD in Figure 2, prisoner I choosing Defect (confessing) is a Nash Equilibrium because opting for the other alternative, Cooperate (not confessing) always gives a worse outcome. This is irrespective of which strategy prisoner II plays. The same argument applies to prisoner II, meaning that the sole pure strategy solution for the game is for both prisoners (players) to play the dominant strategy, Defect, by confessing. What makes the PD such an important experimental tool is that it seems counterintuitive because the strategy that gives the best overall outcome for both players is the Pareto optimal solution where they both cooperate (not confess).

***Advantages of using Prisoner's Dilemma as a method.*** PD game is widely used in Game Theory as it provides a good framework on how to strike a balance between cooperation and competition, which is a very useful tool to analyse strategic decision-making. PD has become a reputable tool for economics and business research especially on strategizing their products. For instance, in the low-cost airline industry in Europe, Ryanair and EasyJet are both airlines selling similar products. Each must decide on the pricing strategy. Either one can exploit the market by selling their air fares as low as possible or as high as possible or else either stay clear from the rivalry by flying into different routes and share the profits from the same pool of customers.

The challenge in this game is for each player to make choices without knowing what the other has decided. Given the uncertainty and the payoff structures, Nash equilibrium is

the decision point, where it is possible for the player to respond following to the opponent choices, despite that it might not be the optimal outcome for both. Yet, it is the most advantageous decision the players can make when neither one of them knows what the others will decide (Gottman, 2011).

Another key strength of the Nash equilibrium is the simplicity of the concept with power to analyse real world events. For instance, in the Cold War, the relationship between the U.S and the USSR were deemed to show the real-world application of the Nash equilibrium (Downs, Rocke & Barsoom, 1996). Both became locked into an equilibrium whereby each side was aware and knew each other's positions but held to the strategy. One of the benefits, to consider the Nash equilibrium is, it creates a basis on how to analyse the role of reward or pay-offs in a social interaction, which can change the order of Pareto optimality. Pareto optimality is best explained by scarcity. Imagine that, there is a short supply of Jelly Babies and the only store that still selling it, start charging customers with a ridiculously high price. Encountering this kind of dilemma in every day lives makes a study of game theory fascinating, especially as it can be applied to real-life events.



## 11.5 Chapter Summary

This chapter offered a general overview of methods in Social Psychology by way of considering how to understand and operationalize the voting dilemma faced by contestants on the WL game-show. The WL is advanced as an important indirect method for studying how individuals make choices regarding whom to choose in a forced-choice setting, from a menu of eight choices (i.e., the other eight contestants). The main reasons that make WL interesting are that, it is fairly easy to see by breaking the game down that it shares some of the common aspects with PD, that are also shared with Milgram's obedience paradigm. WL, Milgram's obedience studies and PD all share common features: A choice dilemma, cooperation/defection and uncertainty. Each approach offers its own unique perspective on these features.

***Prisoner's Dilemma paradigm: Choice dilemma, cooperation/defection and uncertainty.*** PD was described above in terms of its qualities to distinguish between the uncertainty of making payoff-dominated or risk-dominated choice strategies above. PD provides possibly the neatest encapsulation of the choice dilemma, cooperation versus defection and uncertainty. Not only is it a very straight forward dilemma to understand but it is based on a mathematical theory. Game theory can be applied to a wide range of situations and is widely used throughout academia.

***Milgram's paradigm: Choice dilemma, cooperation/defection and uncertainty.*** The choice dilemma faced by participants in Milgram's paradigm became a decision on whether to administer an electric shock. Once this choice had been made the choice

dilemma was repeated again and again until the participant either chose not to administer the electric shock or the shock intensity reached its maximum level. In this fashion, the choice became more of a dilemma as the intensity of the shocks increased. The participant was in some ways lured into the choice dilemma, because following a convivial meet and greet with the *learner* and *experimenter* and having sampled a *test* shock, the participant then began to instigate the shock sequence most likely unaware of the escalation in intensity and potential ramifications of administering very severe shocks. It would be interesting to know if participants had begun at the highest level of shock and worked down to administer as the first shock.

Participants in Milgram's paradigm also face the dilemma of cooperation and defection. With each electric shock, they administer they cooperate with the *experimenter* and defect from the *learner*. One way of thinking about the choice dilemma is where the tipping point is between cooperating with the *teacher* and cooperating with the *learner*. The dependent variable Milgram used was effectively when this switch occurred.

As the intensity levels of the electric shocks increased it is reasonable to assume that the uncertainty participants experienced as they moved to the next level increased. As the uncertainty increased participants experienced stress and agitation regarding their choice dilemma. The uncertainty grew as the participants continued. Participants eventually handled this uncertainty once it hit some trigger point. At the trigger point the participants faced another dilemma of whether to refuse to continue to administer the shocks or accept blithely that they would just continue to the maximum and absolve responsibility to the *experimenter*.

***WL paradigm: Choice dilemma, cooperation/defection and uncertainty.*** This chapter argued that the WL TV game-show provides an alternative method on how social interactions can be observed. This particular TV game-show is unique because it seems to encourage cooperation and defection. Cooperation and coalition were required by contestants working together to build a temporary pot of money for the benefit of all the contestants (i.e., refer to **M1.2**, p. 60 for the WL game-show format). However, the game is structured with a ruthless twist because there can only be a single winner that receives the payoff of the accumulated prize fund at the end of the game. The other contestants are equal “losers” in the context of WL because they each gain nothing.

Contestants defect too, by voting for another contestant to be eliminated from the show in each round as the weakest link. The contestants therefore have to navigate a complex fast-paced scenario of answering quick-fire questions, *banking* and then voting for a peer to be eliminated as the weakest link.

Fevrier and Linemmer (2006) see the WL in terms of Game Theory and suggest that the strategies involved switch between being payoff dominated and risk dominated. In the early, less risky, rounds of WL the payoff incentives dominate when the best strategy is to vote off the weakest players. The final voting round with the three remaining contestants is different though, particularly when there is a single player that is clearly superior. The two weaker players then have an opportunity to “gang-up” to eliminate the strongest player. This is similar to the Stag-Hunt game, where two hunters have to each decide whether to risk joining forces to catch a stag or work alone to catch a guaranteed hare (Osborne, 2009). The stag can only be caught if both hunters work together, but the

hare can be caught by an individual. Hunting the hare carries no risk but hunting the stag requires coordination but gives a better payoff. In terms of WL, the strongest player is indifferent between the two weak players and votes with a probability of 50-50. It then leaves the two weak players in a stag-hunt game. They can be coordinated and catch the stag (i.e., strong player) or be risk averse and vote for the other weak player: if the other player subsequently voted against the strongest player then they would be most likely to be voted off as the “strongest link” who decides whom to be eliminated in the event of a tie.

Fevrier and Linemmer (2006) observed 36 episodes of the WL aired in France. Empirically it appeared that the contestants coordinated on the Nash equilibrium where the strongest-link was not voted out. Interestingly, they found that the strongest link was never eliminated in the first two voting rounds and rarely after that until the final voting round where 5 (14%) were eliminated. The weakest links were eliminated, as their performance would predict, much more frequently with a ratio of about 50% in each round. It still remains the case though, that in about half of the voting rounds the voted weakest was not the actual weakest link (i.e., at least as measured by Fevrier and Linemmer). This suggests either that it was not always very clear who the weakest link was or that the contestants voting was subject to some form of bias.

From a strategic point of view, the contestants’ optimal strategy was to cooperate in the early rounds to accumulate a large cash prize. In the later rounds, there is some incentive for cooperation to reduce and a switch to an incentive to defect. However, Fevrier and Linemmer (2006) found no empirical evidence for this. The final *Head-to-Head* round involved the last two remaining contestants in a pure zero-sum game.

*Social context of data collection.* The WL contestants, although performing in a contrived artificial environment, were all very well informed about the rules of the game and what to expect on the day that they turned up to be a contestant, after all, the contestants made the effort by requesting to take part in the TV show's recording. Therefore, the format of the show and its rules and layout were all very familiar to the contestants. Indeed, considering the WL contestants as participants in a Psychology study they compare very well to the participants turning up to take part in Milgram's paradigm or players recruited to play PD. In these cases, it is generally the case that the participants will not have the familiarity with Milgram's social context or will typically be unfamiliar with PD and its nuances. Therefore, as naive participants to an unfamiliar testing venue without preconceived notions of what to expect, participants' performance will be affected by the novelty of the situation as an extraneous variable on top of the effects attributable to the carefully crafted experimental variables.

WL contestants will also be expected to invest their full attention to the game task at hand as they will be appearing on national TV as well as in front of a TV audience. The stakes are high and not just in monetary terms of the prize. Clearly, they wish to avoid making a fool of themselves in front of friends, family and the nation. Further, the prospect of winning a considerable prize pot is an added incentive. Participants in laboratory studies run by university researchers cannot hope to elicit the same level of attention and commitment from their participants.

The next part in the introduction **I2**, offers an evaluation of the key studies relevant to the thesis based on the WL game-show.

## **Chapter I2:**

### **Evaluation of The Weakest Link TV Game-show**

“In addition, this game-show [WL] provides an ideal laboratory to study human decision-making. The rules are well-defined and the stakes are high, something that is not easy to replicate elsewhere” (Barmish & Boston, 2009, p. 3).

#### **I2.1 Chapter Outline**

This chapter offers a critical evaluation of literature pertinent to considering the voting decisions made by contestants in the WL game-show setting. The argument is made that the WL game-show format offers an excellent resource for studying behaviour. The remainder of this section provides a detailed critical evaluation of the use of analysing TV game-shows as a methodology, with a particular emphasis on weakest link. The WL TV game-show has been used to study a diverse range of issues including discrimination/prejudice, strategic game playing and proximity. These are considered in turn. The chapter finishes with a consideration of the research aims and objectives.

#### I2.1 Chapter Outline

I2.2 Using the Weakest Link game-show to test for discrimination, strategy and proximity.

I2.3 Evaluation of the Weakest Link Game-show Studies

I2.4 Research aims and objectives.

## **I2.2 Using the Weakest Link Game-show to Test for Discrimination, Strategy and Proximity.**

The opening of chapter **I1** alluded to the seeming contradiction that despite being almost constantly embedded within wide ranging social milieus, researchers still find the phenomena of social psychology difficult to pin down and study. This is because human social life is self-evidently very complex and is not experienced by an individual in neatly packaged observable units. The real challenge for many social psychologists then is to somehow operationalize aspects of social behaviour so that they can be more neatly defined, measured and controlled for. Otherwise, it is rarely the case that human social behaviour is presented in such a fixed methodical fashion available for precise scrutiny by an analyst. One such rare exception, however, is the case of the TV game-show, for in this arena people behave according to clear observable rules in a way that is open to outside public surveillance. The WL TV game-show has proven to be one of the most popular amongst researchers for the reasons outlined explained below.

**What WL can tell us about prejudice and discrimination?** The experimental study of prejudice and discrimination is problematic for a number of, mostly obvious, reasons. One of the most notable ways in which the study of the WL has had greatest impact is with respect to testing between theories of discrimination. This section reviews the predominant studies that used WL to explore discrimination. First the academic study of discrimination is presented in terms of theories derived from the fields of Psychology and Economics, definitions of discrimination and problems measuring it.

*Discrimination/prejudice from a social psychology perspective.* “This review argues that research on implicit prejudice, largely developed by Psychologists, provides an important new understanding of the basis of discrimination and should be incorporated in sociological accounts” (Quillan, 2006, p. 299).

There exists an enormous literature on racial and gender discrimination. However, measuring discrimination is a particularly challenging empirical task not least because any form of active discrimination on the grounds of race, gender, nationality or religion is almost universally eschewed (Levitt, 2004; Quillan, 2006). Discriminatory behaviour is outlawed in most international, national and corporate policies. The stigma associated with discrimination poses a particular problem as individuals will naturally avoid labelling themselves as bigots in any kind of conscious self-report study or experimental manipulation of variables associated with gender or race.

One contradiction is that, there is a near unanimous acceptance that discrimination should not be tolerated (Quillan, 2006). However, prejudice and discrimination remain strong. There are some ways to capture this emotion both at an explicit level but also at the implicit level. Some studies test for the existence of discrimination, others test for what kind of discrimination is evident.

Whereas prejudice is a negative attitude about members of a group, discrimination is displayed as a negative act directed towards the members of a group. Discrimination usually develops from prejudice and is associated with negative views and behaviours. A major development in social psychology occurred with models of Implicit Social



Cognition. These models showed that the attitudes and stereotypes that people adhere to need not be conscious and reasoned. Instead, attitude, stereotypes and discrimination were potentially more likely to be unconscious and implicit. Greenwald and Banaji (1995) provided useful operational accounts for implicit attitudes and stereotypes.

“Implicit attitudes are introspectively unidentified (or inaccurately identified) traces of past experience that mediate favourable or unfavourable feeling, thought, or action toward social objects...Implicit stereotypes are the introspectively unidentified (or inaccurately identified) traces of past experience that mediate attributions of qualities to members of a social category” (Greenwald & Banaji, 1995, pp. 8 & 15).

Therefore, discrimination, from a psychological perspective at least can be both explicit but potentially more often implicit. In explicit prejudice a person is fully aware and conscious of their prejudice and the discrimination that follows is overt, direct and explicit. Even when a person is explicitly prejudiced, they might not necessarily go on to enact their unjustified attitudes in the form of discrimination. However, it can also be the case that other persons are unaware of their own bias and unconscious of the discriminatory behaviour it leads to. For these individuals the prejudice and discrimination is both automatic and implicit.

A meta-analysis of studies investigating the link between prejudice and explicit discrimination (e.g., Greenwald, Poehlman, Uhlmann & Banaji, 2009) found a reliable if weak relationship. The weak relationship is understandable given the social desirability of not wishing to appear discriminatory. Studies of implicit prejudice are potentially more

revealing because they get around the problem of participants recognising and conforming to socially desirable responses. In a meta-analysis of implicit prejudice/stereotypes and their link with discrimination a weak to modest reliable relationship existed but one that was stronger than that found for explicit measures (Greenwald et al, 2009).

Many methods have been used to demonstrate implicit social cognition. These range from mere exposure, Implicit Activation Tests (IAT), stem completion tests and have been used in fields as diverse as attitude formation and memory function. Many social psychology phenomena are associated with Implicit Social Cognition: halo effect, attribution theory, fundamental, disposition, situational attribution effects, self-esteem, attitude formation, stereotyping (i.e., including gender and race).

*Implicit race and gender stereotypes.* Many studies have shown that there are likely to be deeply held automatic, unconscious stereotypes of race and gender. Although Gaertner and McLaughlin (1983) found there were no differences in reaction times when white respondents had to judge if lists of negative traits were related more to whites and/or blacks, when the word lists were positive (e.g., White clever vs. Black clever), white respondents were quicker linking them to white targets than black targets. The implication behind these indirect measures of implicit social cognition is that the connections are formed more quickly between concepts that have already been formed. Therefore, the prejudice was pre-formed.

White participants, surreptitiously primed with terms that were mostly stereotypically associated with African-Americans, were affected on their judgements on

subsequent tasks (Devine, 1989). Judgments made in a seemingly race neutral setting still attributed male targets as significantly more aggressive than controls exposed to non-stereotypical terms. In a similar fashion to Gaertner and McLaughlin (1983), this result highlights the importance of implicit social cognition and the role of automatic cognitive processes underlying stereotyping and prejudice in implicit racism.

Dovidio and Gaertner (2004) developed a theory of aversive racism based on dissonance occurring between the feelings of many white Americans with liberal views yet still harbouring negative stereotypes and attitudes about Blacks. An unusual conflict emerges where the individual at once conforms to the anti-racist political agenda but still holds negative attitudes towards blacks. Such automatic application of stereotypes forms the foundation for implicit prejudice. Word, Zanna and Cooper (1974) found that white interviewers discriminated against black applicants for a job more than white interviewees. Further, when these kinds of discriminatory non-verbal cues were later actively introduced into the interviewer's repertoire it was found that white interviewees performed worse.

Implicit Gender Stereotyping occurs in a similar way to race stereotyping. Another system justification theory of discrimination (Jost, Banaji & Nosek, 2004) suggests that discrimination is justified based on the status quo. It is the way of the world that those in the in-group benefit from the advantages they enjoy and the out-group members are in that position for a reason. This can work both ways with the out-group almost falling into the trap of going along with the system justification theory. One possible example is the bias shown in self estimations of IQ. Women typically judged their own IQs significantly lower relative to males across many variants of participant groups. This effect might be in

part a result of male hubris and female humility or even one where the system justification theory maintains gendered stereotypes relating intelligence as a stereotypically male trait (for review see Furnham, Kidwai & Thomas, 2001). It is not surprising in view of this kind of research that long standing discrimination occur in grading essays higher with male as opposed to female names (Goldberg, 1968). This serves to highlight the implications for this kind of subtle bias can have by becoming widely held stereotypes held throughout a culture operating again as an unconscious automatic bias. This reinforces again the importance of anonymous marking in examinations. Meta-analysis of similar studies support Goldberg's finding (Banaji & Greenwald, 1994).

The cognitive process that underpin implicit social cognition can be allied with the *fast* and *slow* systems reviewed by Kahneman (2011). The System 1. Thinking is fast, parallel, automatic and unconscious akin to the implicit prejudice and discrimination. The explicit prejudice and exhibition of discrimination fits with the System 2 way of thinking, slow and deliberate but conscious and overt.

*Expressions of discrimination – institutional/systemic prejudice/discrimination.*

Expressions of discrimination can occur in a variety of forms including hostility (i.e., verbal and non-verbal), aggression, avoidance and the denial of opportunities and equality. Even organizations or other formal groups can become prejudiced and display discriminatory practices. Institutional or systemic discrimination occurs when the policies or practices of social bodies leads wittingly or otherwise to negative consequences for members of certain groups. Ethnic minorities face worse health, school performance and

stiffer sentences. The “ethnic penalty” refers to the disadvantage incurred by ethnic minorities in the job market (Heath and McMahon, 2005).

*An example of institutional racism.* One fairly recent high profile example in the UK arose from the MacPherson report published on 24 February 1999, on the inquiry by the Metropolitan Police Service into the murder of the black teenager Stephen Lawrence. The report concluded that the police investigation was “marred by a combination of professional incompetence, institutional racism and a failure of leadership by senior officers” (Home Office, 1999, para 46.1) and went on ‘...institutional racism affects the MPS (Metropolitan Police Service), and police services elsewhere.’ (ibid, para 46.27). The report made 70 recommendations designed to achieve ‘the elimination of racist prejudice and disadvantage and the demonstration of fairness in all aspects of policing.’ (ibid, recommendation 2).

*Definitions of discrimination.* One of the most contentious aspects of the Macpherson report was its branding of the Metropolitan Police Service as institutionally racist, something the then Chief Constable refuted. More recently, Sir Bernard Hogan-Howe has nevertheless conceded that, “If other people think we are institutionally racist, then we are. It is no good me saying we are not and saying you must believe me. [That would be] a nonsense, if they believe that” (The Guardian, 5<sup>th</sup> June 2015). Macpherson carefully defined institutional racism in his report as something that leads to discrimination:

The collective failure of an organisation to provide an appropriate and professional service to people because of their colour, culture, or ethnic origin. It can be seen or detected in processes, attitudes and behaviour which amount to discrimination through unwitting prejudice, ignorance, thoughtlessness and racist stereotyping which disadvantage minority ethnic people. (Home Office, 1999, para 6.34).

Another useful description of discrimination is presented more generally as “...behaviour directed towards category members that is consequential for their outcomes and that is directed towards them not because of any particular deservingness or reciprocity, but simply because they happen to be members of that category” (Correll, Judd, Park & Wittenbrink, 2010, p. 46). Taken together, these formulations of discrimination include a prejudice that can be unwitting or implicit and a behaviour that has perceived negative connotations. Furthermore, these can exist at the level of the individual and the corporation. Given that an integral feature of the WL game-show involves the contestants voting for one another and that the vote can be considered as doing something bad or nasty, then it becomes easy to see that analysing how contestants distribute their votes could be used to find bias and test if contestants potentially harbour prejudice and demonstrate discrimination (as endogenous or endogenous factors). This could be done by simply counting whether the vote frequencies against various groups or minorities departed significantly from chance expectations.

***Discrimination/prejudice-from a behavioural economics/economics perspective.***

The psychological theories of discrimination described above are largely concerned with how a social actor comes to form prejudiced views and acts in such a way to disadvantage

a target group or its members. These accounts focus on the formation of attitudes, displays of behaviour and cognitive processes underpinning explicit and implicit discrimination. Theories derived from the field of economics consider the economic actor and arrive at two main accounts for discrimination, one taste based and the other information based. Whereas the psychological theories seem to fit more with the *fast* intuitive System 1 thinking, the economic perspective possibly takes the more rational, deliberate, utilitarian System 2 way of thinking.

*Taste based models of discrimination (Endogenous factor).* Becker (1957) suggested that the source of discrimination lay with the discriminator and their active considered choice to target a specific group. This taste-based, or preference-based discrimination occurs “...because people simply do not like members of certain groups” (Antonovics, Arcidiacono & Walsh, 2005, p. 919). If members of group *A* have a taste-based dislike for *B*, then they should exhibit that discrimination throughout all voting rounds according to Levitt (2004) although Antonovics et al. (2005) explain that the discrimination will wane in later rounds as the likelihood increases that remaining players will be the ultimate victor. Then it is less likely that they will maintain prejudiced taste-based negative voting that could potentially jeopardies their own future chance of success.

*Information based models of discrimination (Exogenous factor).* Information based models suggest that the source of discrimination lies in mistaken negative beliefs attributed to groups and their members. Whereas taste-based models incorporate some form of “dislike” or “animus” towards the out group, under the information-based model there is no emotional force required for the discrimination to occur. In information-based or

statistical-based discrimination occurs because “...people use group identity as a proxy for unobserved ability” (Antonovics et al., 2005, p. 919).

Statistical discrimination as the exogenous factor can take two forms using correct priors (Antonovics et al., 2005). One form predicts that one group, say  $A$ , perform consistently worse than another group,  $B$ . In the context of WL this would be by doing something like answering questions inaccurately. Another type of statistical discrimination is that information regarding the ability of one group is clearer than for the other group. This implies that in-group performance is weighted more heavily than out-group performance. Hence, poor performance is weighted more heavily when it is by a member of one's own group.

If group based discrimination occurs, it should be evident in the pattern of voting. Antonovics et al. (2005) explained it in hypothetical terms for groups  $A$  and  $B$ , under two different scenarios, one where the ability level between the groups differed and another where the quality of information between the groups differed. In the former, where  $A$  plays the game better than  $B$ , if all players are capable of interpreting this then it should be the case that all contestants will vote against members of  $B$ . This shows as a statistical based discrimination where both groups discriminate against  $B$ . However, when  $A$  and  $B$  play the game equally well, but members of  $A$  have less accurate information about the skill level of  $B$ , then  $A$  will be more likely to vote against other members of  $A$  because skill level is equal but there is imbalanced information signalled about the two groups. Conversely, if voting in WL was for the best player, the strongest link, then the predicted pattern would shift and  $A$  would be more likely to vote for  $A$  under the same



circumstances. Another possibility is that  $A$  discriminates against  $B$  based on mistaken prior information. Then  $A$  target  $B$  until the prior is corrected.

***What studying WL tells us about discrimination?*** Two studies dominate the literature on the use of WL to test theories of discrimination in the economic literature, both of which used contestants voting patterns to measure the extent of discrimination and its type, taste based vs information based (Antonovics, et al., 2005; Levitt, 2004). These are reviewed in turn.

*Levitt (2004).* Levitt observed the pattern of voting in all the voting rounds in 161 episodes of the U.S version of WL TV game-show (25 prime-time and 136 daytime shows;  $N = 1,016$  contestants) to test two ideas (a) the extent of discrimination and (b) to distinguish between taste-based and information-based discrimination. The prime-time show comprised six voting rounds and the daily show had four. A regression model was used with the votes received by a contestant as the dependent variable and independent variables of contestant demographics (i.e., race, age, gender, educational level, region of residence), and performance regarding answering questions et cetera. Early, middle and late voting rounds were analysed separately.

The premise behind measuring the first aim, (a) the extent of discrimination, was fairly straightforward. A vote in the context of WL was defined as a bad thing. It follows therefore that if contestants harboured prejudice against a group, so that it spilt over into active discrimination against its members, then in the context of WL that group would be expected to receive significantly more votes. To fit with the definitions above

discrimination is considered as an unjustified negative act (i.e., weakest link vote) aimed at the members of a group based purely on their being perceived as part of that group. In other respects, they are undeserving of the negative act. Levitt (2004) recorded the age, gender, race, education, employment status for each contestant. With regard to gender, there was little difference between votes received by males and females with female receiving slightly fewer votes. On closer analysis however, there was evidence of in-group favouritism with females more likely to vote for males and males more likely to vote for females. With respect to race, blacks received slightly more votes in the early rounds than whites but they received slightly fewer votes in later rounds than whites. Therefore, there was no evidence of gender or racial discrimination (black or white) in the study.

To test the second aim (b) to distinguish between taste-based and information-based models of discrimination, Levitt (2004) adopted the reasoning that contestants should adopt a switch in voting strategy as the game develops. In the first voting round, all the contestants are required to cast a vote and there are many contestants to choose from, but following successive eliminations the final voting round only had the last three remaining contestants. In the early rounds, it makes strategic sense to vote off the “weakest” contestants to help generate a larger pot of money, but in the final voting round it is strategically beneficial to vote off the perceived “strongest” player to ensure that in the final *Head-to-Head* the contestant is facing the weaker of the two potential opponents. Strategic considerations aside, if contestants are prejudiced and engaging in taste-based discrimination against a group’s members then that group should expect to receive more votes across all the voting rounds. However, if the discrimination is information based, the group should expect to receive significantly more votes in the early rounds but

significantly fewer votes in the later rounds. This is because taste-based discrimination is targeted at the group and its members which includes animus on the part of the discriminator towards the group. Information-based discrimination is different and more mercenary, here the discriminator discriminates on the basis of a negative stereotype about a group and its members. For example, a group might be targeted based on perceived low skill levels, then the members of this group will receive significantly more votes in the early rounds but significantly fewer at the end. Levitt's appeal for using WL is based on this unique feature of the game-show. It is rare that such distinctions can be made between the predictions from these different forms of discrimination using other empirical methods or data in real markets. Levitt found that whereas there was no evidence for female/male, and black/white discrimination, but there was evidence for taste-based ageism and information-based discrimination targeted towards Hispanic contestants. Contestants over the age of 50 years old received significantly more votes across all voting rounds and Hispanics received significantly more votes in the early rounds and significantly fewer votes in the final rounds.

*Antonovics, Arcidiacono and Walsh (2005).* Antonovics, et al., 2005 also observed all the voting rounds in the U.S version of WL TV game-show (103 episodes; 28 weekly and 75 daily shows;  $N = 682$ ), with a similar aim to Levitt (2004) of testing between taste- and information-based theories of discrimination, that they refer to as preference and statistical-based discrimination. Three demographic groups were studied: men, women and whites. Their voting behaviour was measured as a fraction termed the Group Bias Statistic (GBS). The numerator captured the voting of an individual,  $i$ , by assigning the value of 1, if voting for a person of the same group in the same round (0 if voting for

someone in another group). The denominator was given by the ratio  $\frac{G-1}{N-1}$ , where  $G$  is the number of  $i$ 's type in the round and  $N$  is the total number of contestants in the round.

Consider, an example of how GBS might work by taking  $i$  to be a female contestant voting in a round made up of 5 women and 4 men, (giving  $G=5$ ;  $N=9$ ) then, when  $i$  voted for another woman  $GBS = 2$ , otherwise, if  $i$  votes for the out-group male,  $GBS = 0$ . If the voting is biased regarding gender such that all of the women in that round each voted for another woman then their combined GBS would sum to 10 ( $5*2$ ), with a mean of 2, but if they all voted for men then the GBS sums to zero, giving a mean of zero. Over the course of many rounds, if there is no bias between voting for the in-group women, or the out-group men, then the GBS has a mean of 1, indicating no discrimination. A value greater than one indicates discrimination against one's own group, the in-group, and a value less than one points to discrimination against the out-group.

The main finding was a significant discrimination by females voting for male contestants as the weakest link significantly more than other females. Importantly, there were no significant race or gender differences between the contestants' ability to answer the questions, so it seems likely that women had a preference-based discrimination against men. The GBS of 0.739 and 0.783 were found for the first-round of voting by females on males in the daily WL, and weekly prime-time WL shows respectively. Both were significantly less than the predicted *no discrimination* measure of 1. A significant GBS of 0.783 was also found for second-round voting in the daily show. Antonovics et al. (2005) argue that finding evidence of discrimination is stronger when it appears in earlier, rather than the later voting rounds, because there is a greater pool of voters, with no history of prior voting untainted by performance in the previous rounds. Males showed significant

discrimination against women in the third voting rounds of both the daily and weekly shows, although it is unlikely that these observations have any deeper ramifications as there is little strategic reason why discrimination should be restricted to third round voting. One possible reason for males' round three voting discrimination against females is a reaction to the recognition that they (i.e., males) have mainly been targeted by women in the first two voting rounds.

Estimated conditional logits, by show and round, were also used to support the findings above. Estimations were based using the chance of a voter picking another contestant as the *weakest link* based on their characteristics. The conditional logit allowed for the considerations of interactions between contestants' characteristics. The main analysis was mostly relevant for the first and second voting rounds, for reasons already stated above. A large effect was found as women were again shown to be more likely to cast votes against men rather than other women, an effect that was immediately lost after the first-round. When women voted in the first-round, a particular male had a 23.3% likelihood of being the recipient of a vote relative to 15.1% for a female recipient.

Antonovics et al. (2005) entertain three possible explanations for why women vote against men. One is that strategic alliances could be formed where the members collude in a joint venture to target another group. This could be explicit and agreed upon or implicit and tacit. In either case, it should be possible to explore collusion through contestants' voting patterns. Explicit collusion by women targeting males was discounted because the total votes cast by other women had no more predictive power than the total votes cast by men in accounting for females voting in the first-round. Implicit collusion was also

discounted because if it were occurring then after the first-round it should become even more prominent in the second and third rounds, and so on until all males had been eliminated. This was not the case.

Statistical-based discrimination was also ruled out because an analysis of game performance shows negligible differences between the three groups for answering questions correctly in the first-round: male 67.7, female 66.8%, black 63.9%, white 67.5%. There was no indication of statistical discrimination regarding ethnicity or gender. Furthermore, if there were some other aspect of poor performance not captured by question accuracy, (e.g., hesitancy, lack of confidence) then males would also be expected to vote against males. This result is important because although there might be the basis for statistical-based discrimination in the labour market relating to women it is unlikely to be present in WL. The only remaining explanation was that women preferred playing the game with other women.

**What studying Weakest Link game-show tells us about strategic game playing?** The WL involves interesting questions of strategy, both in when to bank and whom to vote off, that have attracted interest from various disciplines. The review of the studies above based on discrimination relied mainly on the aspect of the WL game-show that made use of the voting decisions made by contestants at the end of each round. However, another class of studies exist that use the WL game-show as a means to investigate strategic aspects of game playing from a Game Theory perspective.

It has not escaped the notice of researchers that TV quiz shows like WL make ideal observational field experiments because they comprise the key ingredients of game theory: a finite group of players must select from a fixed set of actions to play for well-defined payoffs. For example, WL has been used to assess the optimal banking strategy in economic decision-making (Haan, Los & Riyanto, 2011), the trade-off between risk and return strategies in game playing (Barmish & Boston, 2009) and the optimal voting strategy to adopt (Février & Linnemer, 2006). Perhaps the most relevant for this thesis relates to the role of strategy in mixed-gender group interactions as endogenous and exogenous factors for the voting bias (Valenzuela & Raghurir, 2007).

Valenzuela and Raghurir (2007) observed 20 episodes of the U.S version of WL finding that, there seems to be a skewed gender effect different to those above in that females were frequently finalists but tended to lose in the final. They tested the hypothesis that female contestants are favoured, by not being voted off in the early rounds, but then targeted in the later rounds. They tested this with a simulation of WL, predicting that women would be more likely to be retained in the group than men but they would be no more likely to be the eventual winner than a man. They observed the performance of contestants and computed the *measure weakest link* as well as recording the *actual weakest link* as voted by the contestants, noting that in 67 out of 120 rounds the *actual weakest link* was not what they had computed as the *measure weakest link*. Furthermore, in the rounds when the contestant voted weakest link was not the *measure weakest link*, the *measure weakest link* was female significantly more often than male (64.18% to 35.82%). Voting the weakest link was not therefore gender neutral. Their results mirrored in part the observations of Levitt (2004) and Antonovics et al. (2005) above.

Haan et al. (2011) analysed the banking strategies of contestants and found that the contestants were not efficient and used non-optimal strategies to make their banking decisions. Effectively, they used a bounded rationality model to address the issue of when it is the best time (after a sequence of how many questions) for a contestant to *bank* the money. Following this they found that contestants banked too early and were therefore not rational. Barmish and Boston, (2009) took this question further and modelled game playing by comparing theoretically derived gaming strategies with those actually used by the contestants. They suggested that strategies were not simply based on maximising the payoffs in the game, by developing the most efficient ‘banking’ strategy, but were also mitigated by taking into account risk in banking and voting considerations. Once they modelled behaviour taking into account the dual concerns of risk-return, then the models fit the contestants game playing in a rational and efficient manner.

As such, in making a voting choice, contestants might be expected to avoid conflict or risk as proposed in prospect theory whereby, choices were made based on the probabilistic alternative where the individual made the decision based on known outcomes (Kahneman & Tversky, 1979). According to Hertwig (2012), it was found that a description-experience gap would lead to different choices particularly in rare and high consequence events. In addition, the nature, degree and perceived risk are also the vital elements in preference formation with or without prior knowledge which affect the decision-making process (Pras & Summers, 1978; Sharifpour & Walters, 2014; Simcock, Sudbury & Wright, 2006).



**What studying WL tells us about the effects of proximity? *Proximity, the Law of Attraction.*** Individuals form quick temporary interpersonal relationships towards others that are close to them simply on the basis of their physical proximity. If continued, this proximity can lead to stable friendship networks. This proximity principle was first illustrated by Newcomb (1956) in his observations of shared accommodation among transferred students, who were initially strangers but developed friendships due to their closeness and continuous interaction. In addition, Hall (1963) developed the area of proxemics and suggested that, the interpersonal distances between individuals can be described by different zones of acceptability based on the relationship with the individual (i.e., intimate, personal, social and public). The interaction is based on each individual as the nucleus of the interaction and different implicit zones of acceptable space are built around them. These socially acceptable physical zones vary dramatically according to different cultural norms and custom.

In the WL game-show, contestants interact in a carefully constructed spatial configuration where the spatial relationships between contestants are clearly defined. Thus, the proximity of neighbours can develop a special camaraderie among contestants. Matthews and Matlock (2011) suggested that there was a relationship between social distance and physical distance. In their exploratory studies, they explored an association between how individuals conceptualized relationship and space. They illustrated it in a form of narrative on how do people associate their thoughts and relationship.

Conversely, Trope and Liberman (2010) provided a different way of understanding social proximity. They suggested a Construal Level theory (CLT). This CLT used

different *psychological constructions of distance* based on either spatial, temporal or hypothetical proximity. More proximal events become therefore more concrete whilst the more distant are considered weaker and more abstract. CLT can be divided into two different levels: high-level occurs when there was a significant psychological distance and the perceived information was comprehensive (abstract), low-level thinking was perceived to be detailed-specific (concrete) when there was close proximate interaction.

***Proximity and doing bad things: Milgram (1963) and Goddard, Ashley, Fuller & Hudson (2011).*** The spatial relationships between actors in social situations can have a profound impact on their subsequent social behaviour. For instance, there is a clear link between the proximity of the *learner*, *teacher* and authoritarian *experimenter* in Milgram's (1963) classic obedience paradigm, and the level of obedience elicited in the *teacher*. Initially, obedience was at its greatest when the *teacher* was spatially remote from the *learner*. Issuing instructions from a separate room seemed to make it easier for the *teacher* to administer "punishment", but, in replications where the *learner* could be seen by the *teacher*, obedience reduced until it almost disappeared when the *teacher* was instructed to physically place the *learner's* arm onto an electric plate in order to receive the punishment of an electric shock. Similarly, the *teacher's* level of obedience to authority diminished as the distance between the *teacher* and the authority figure of the *experimenter* increased (Milgram, 1974). Milgram's study has some parallels with WL because the participants/contestants had to do something nasty or bad to another person. The dilemma for Milgram's *teacher* was how far on a scale of punishment, voltage of electric shocks, a participant was prepared to go.

The dilemma for the WL contestant was who to choose to receive the *bad thing*. Goddard et al. (2011) reasoned in simple terms that one factor that could influence this choice is the physical proximity of the recipient. In the same way that *teachers* become more reluctant to inflict harm the closer they get to their victims so the WL contestant might find it uncomfortable inflicting harm on those closest to them. Given the fixed arrangements of contestants on the WL game-show, it was relatively straightforward to measure how often the *voter* (i.e., contestant who nominate other contestant) and *votee* (i.e., contestant who received the nomination from the *voter*) were direct-neighbours, or next-door-but-one neighbours, or next-door-but-two neighbours and so on. It was also easy to work out how many votes would be expected between *voter* and *votee* when they were direct-neighbours, or next-door-but-one neighbours, or next-door-but-two neighbours and so on. This was derived based on simple probability alone.

Imagine a model where the nine contestants were only capable of making a random vote at the end of each the WL round. Whereas the WL was used to test discrimination and strategy in the sections above, envisaging the contestants as social actors full of discrimination and bias, or strategic agents acting rationally to maximise utility, Goddard et al. (2011), treat contestants as naive automata and predict their voting behaviour based solely on probability. Goddard et al. (2011) make the assumption that even if bias does exist in WL, in whatever form, then it would be evenly distributed across many episodes and cancel out. If a systematic bias existed however, then it would be amplified over the course of many shows. If proximity effects voting behaviour, then a simple linear function between vote frequency and proximity would be seen such that there would fewer votes

than expected when *voter-votee* were proximal but more than expected when *voter-votee* were distant.

The voting patterns of contestants in 72 episodes of the UK version of the WL were observed. Goddard et al. (2011) found that there was an effect of *voter-votee* proximity, but not one that conformed to the simple linear prediction above. Instead *voters* showed a dramatic *neighbour effect* (NE) by significantly avoiding choosing their direct-neighbours. However, they also voted for contestants furthest away with the predicted expected frequency, therefore showing no bias. The contestants in the middle however received many more votes than expected. Therefore, the results showed a strong effect for proximity but not a simple function *voter-votee* distance. They referred to this as a “neighbour-avoidance” effect but offered no explanation for the non-linear pattern of voting with distance. It is almost as if the voting pattern showed three distinct phases: A *proximal voter-votee* phase where *voters* significantly avoid voting *votees*, a *medial* phase where *voters* significantly targeted *votees* and a *distal voter-votee* phase where the *voter* is unbiased and neutral making the expected number of votes for contestants furthest away.

Goddard et al. (2011) made a further prediction regarding proximity based on the decision-making capabilities of the contestants. They suggested that there were two primary sources of information available to contestants to help them arrive at their voting decision. The first kind of information TYPE I is open, public and overt and would typically be based on the observable game playing ability of the other contestants. For example, consider the game scenario where all the contestants answer every question correctly, bar one contestant that answers all of the questions incorrectly. In this scenario,

they argued that the decision for the contestants regarding whom to vote is easy and straight forward, they would all vote for the contestant that made all of the errors. The contestant would be expected to receive all eight votes (not nine because a contestant cannot vote for themselves). Therefore, in high consensus games like this, the identity of the weakest link is usually clear because they are distinctive for having made mistakes and these judgements can be made based solely on TYPE I information. However, consider contrasting cases where there is limited availability of TYPE I information. Imagine, the situation where all the contestants get all the questions correct, or incorrect, in which case there is limited TYPE I information available with which to make a clear voting decision. In these cases, where there is a great deal of uncertainty regarding the identity of the weakest link, another source of information has to be used to make their decision. This TYPE II information is implicit, private and probably unconscious. Contestants have to guess or use hunches to make their decision as to the identity of the weakest link. In these episodes, it is likely that this *uncertainty* creates low consensus by the voters as to the identity of the weakest link. Goddard et al. consider that individuals' bias and prejudice come to the fore in these situations, including their bias regarding proximity.

The key variable then is uncertainty. Goddard et al. (2011) maintain that the voting decisions in cases of great uncertainty will be based more on prejudice and proximity. Therefore, they predict that as the uncertainty regarding the WL decreases, measured by consensus, the greater the proximity bias will become. They tested this and partitioned the 72 episodes according to consensus of weakest link votes and discovered as predicted that as consensus reduced, proximity bias increased. They concluded that this provided a powerful demonstration of the unconscious influence of proximity on decision-making.

### **I2.3 Evaluation of the Weakest Link Game-show Studies**

Levitt (2004) and Antonovics et al. (2005) carried out studies in parallel to assess discrimination in WL. Although they used different analyses, they largely agreed that there was no evidence for white-on-black or male-on-female discrimination in voting patterns. The discrimination that did emerge was somewhat unexpected because females showed a taste-based/preference-based discrimination against men. It remained uncertain though as to whether women were making an approval vote for other women or a disapproval vote towards men or some combination thereof. In either event, it seemed that the preference/prejudice could not be accounted for by recourse to how accurately women and men answered questions so it would appear that, the discrimination was taste-based as opposed to information-based.

A key feature that both studies alluded to was the recognition that early voting rounds in WL might require different strategic considerations than the later rounds. In the earlier rounds, it was to the advantage of contestants to vote off the weak players, but in the final rounds it might be better to collude to oust the strongest player. It is this shift in incentives that Levitt (2004) claimed was the real academic appeal of the WL because it gives researchers grounds for distinguishing between taste-based and information-based forms of discrimination.

Antonovics et al. (2005), were more circumspect about there being a complete switch in incentives and strategy, but readily acknowledge that the strength of the incentives changes as the rounds progress. They disagree somewhat about the predictions

regarding discrimination. Levitt (2004) maintained that taste-based discrimination would persist throughout all rounds whereas for Antonovics et al., preference-based discrimination would be strongest in the earlier rounds because in later rounds the importance of voting strategically outweighs any personal bias. With information-based, statistical-based discrimination Levitt argued that weak players should be targeted in early rounds but in later rounds they were favoured as it is better to face them in the final *Head-to-Head* than a strong player, hence the switch in strategy. Recall too, however, Février and Linnemer (2006) found no compelling significant evidence of the strongest link being targeted in the final voting round, albeit on a limited sample (20 episodes).

Goddard et al. (2011) ignored any potential attitudinal bias of contestants and instead tested for a simple bias regarding proximity. Their methods differed slightly from those of Antonovics et al. (2005) and Levitt (2004) who used the GBS, conditional logits and regression analysis to model voting behaviour.

Instead, Goddard et al. (2011) simply counted votes according to the relative positions of contestants and compared this count with the expected count derived using basic probability theory. This method of analysis was similar to the group bias statistic of Antonovics et al. (2005) that in turn gives a probability measure of bias. However, the GBS could not be applied by Goddard et al. because even though they recorded contestants' gender they did not identify groups in their analysis of proximity. In line with the work of Levitt (2004) and Antonovics et al., it would be interesting to test if a similar gender effect arises in the UK version of WL.

Levitt (2004) and Antonovics et al. (2005), both made use of all rounds in their analyses and argued that incentives/strategy changed by round, Goddard et al. (2011) meanwhile only measured voting performance in the first-round but argued that grouping episodes according to consensus of group vote for the weakest link could be used as a more meaningful measure of discrimination. Goddard et al. (2011) only used first-round voting because this is the only round where contestants were truly naive, making use of only two types of information (TYPE I & TYPE II).

Another type of information available to contestants and not covered by Goddard et al. (2011) was their strategic voting. This means that, particularly for voting behaviour beyond the first-round, another source of information, TYPE III strategic voting, was brought into the voting decision to be made in subsequent rounds. This could be in terms of “punishment” that Levitt (2004) found where a surviving *votee* from the first-round targets the *voter* in the second-round as a kind of revenge (see next chapter too). Principally though as contestants get voted off and eliminated it becomes more complex to analyse voting as function of *voter-votee* spacing once gaps emerge in the configuration of contestants. Although Goddard et al. lose some explanatory power by restricting their analysis to a single round, they pointed out that more can be gained from looking at consensus as an explanatory variable. Their argument is that bias and discrimination become more prominent as consensus drops. The prediction follows that if Levitt and Antonovics et al. (2005) similarly separated their data by consensus in the first-round then they might be able to test their first aim, to measure the extent of discrimination too. Antonovics et al. actually alluded to this when they noted that with respect to the money banked per round, discrimination increased when the money banked was either very low,



in the lowest quartile, or very high, in the highest quartile. As Goddard et al. point out, the conditions that maximally favour discrimination occurred when the uncertainty regarding the weakest link was greatest, triggering recourse to fallible TYPE II information. This would be expected at the extremes of performance where there were episodes with all contestants performing consistently at one extreme level, for instance where all contestants were strong so there was no distinctive weakest player to vote for. Alternatively, when many of the contestants were poor performers, the contestants had several candidates to choose from. Both scenarios create uncertainty. Levitt predicts that taste-based discrimination would occur in all rounds, Antonovics et al. predict that it is strongest in the first-round and wanes thereafter whereas Goddard et al. predict that it is strongest not only in the first-round, but in those episodes in the first-round where the game information is ambiguous and uncertain, such that, voting bias increased as uncertainty increased as consensus decreased.

Hence, the following section generates the research aim and objectives for this thesis after the evaluation of methods and understanding of the variables exists particularly in the voting behaviour.

## I2.4 Research Aim and Objectives

**Research Aims.** The main aim of this thesis was to explore biases in decision-making and voting with regard to the endogenous and exogenous factors underlying how people made voting decisions. This was done indirectly by observing the voting decisions made by contestants on a TV game-show.

**Research Objectives.** Several hypotheses were developed throughout the thesis and these are extrapolated in the appropriate results chapters that come later. To begin with there were two principal objectives (1. and 2.) and two subsidiary objectives (3. and 4.) that can be broadly stated thus;

Objectives;

1. To demonstrate/re-validate/replicate/extend the original observations of the *neighbour effect* using observations of the Weakest Link TV game-show.
2. To test the validity and reliability of the observations from 1. by finding other scenarios where the *neighbour effect* can be tested.
3. To test whether *neighbour effect* occurs with positive votes (as well as negative votes)
4. To assess *neighbour effect* on tasks involving strategy (e.g., Prisoner's Dilemma)

## **Chapter 2**

### **Methods**

The **Methods** section comprises three main parts (**M1.2**, **M1.3** and **M1.4**) to present the different methodologies used to collect the quantitative data for this thesis. The first part describes the observational field studies on the Weakest Link TV game-show. The second part describes the pilot studies conducted as an attempt to replicate the Weakest Link TV game-show. The final part details the observational field studies using two experiments in the setting of a University lecture.

# **Chapter M1:**

## **General Methods**

### **M1.1 Chapter Outline**

Two different methods were primarily used in the thesis. One of the main aims of this research was to explore how people made voting decisions. This was done in the first instance by using observational field studies of a very specific social context; the WL TV game-show. The second method used field experiments to test whether the findings from the observational field studies could be generalised to other scenarios outside the TV studio.

In each of the results chapters that follow this chapter a brief overview of the relevant method is given and pertinent details relating to the particular study.

M1.1 Chapter Outline

M1.2 Observational Field Study on the Weakest Link TV Game-show

M1.3 Pilot Studies

M1.4 Lecture based Field Experiments

## M1.2 Observational Field Study on the Weakest Link TV Game-show

The first method used in this thesis was an observation of a TV game-show. This section explained the WL TV game-show format and the data collection procedure.

**Rules of the game.** The UK daytime version of the WL game-show was made up of nine contestants that answered eight rounds of general knowledge questions to build up a cash prize. After each round of questions one contestant was voted off until only two remained. They played a final round of questions before going on to compete in a *Head-to-Head* penalty shoot-out style series of five questions each, to determine the eventual winner. This single winner received the cash prize built up by the contestants over the preceding rounds. The other eight players received nothing.

The team of nine contestants stood side-by-side in a semi-circle and took turns answering general knowledge questions from the host who was in the centre. The aim of the WL TV game-show was to create a *chain* of consecutive correct answers to general knowledge question so that the contestants acting as a team built up a cash prize in a single collective *pot*. Table 1 shows how the *chain* built up the prize money in the sequence from £20 for the first correct answer in a sequence to the maximum £1000 for the eighth correct answer in the sequence.

A single incorrect answer broke the *chain* and the *pot* reset to zero and a new *chain* started. However, a contestant could safely *bank* the money built up in the chain by saying “Bank!” on their turn before the question was asked. This money was safe and set aside.

Once the money in the chain had been banked a new chain began from zero. This carried on until the three-minute time limit had elapsed or £1000 had been won in that round (i.e., the time limit gets 10 seconds lower after each round). In addition, any prize money not banked prior to the allotted time for the round was lost.

The question round ended with the elimination of one of the contestants, the so-called *weakest link*. Each contestant was required to vote one player as the weakest link. Voting was done privately by contestants by writing the name of their choice on a board unseen. Once all the votes had been cast the contestants in turn publicly displayed their board and said the name of the contestant they had nominated as weakest link. The contestant with the most votes was the weakest link and eliminated from the game. Ties were resolved by the so-called strongest link, the contestant deemed by the game-show's producers as the strongest player in the round, picking the weakest link.

Table 1

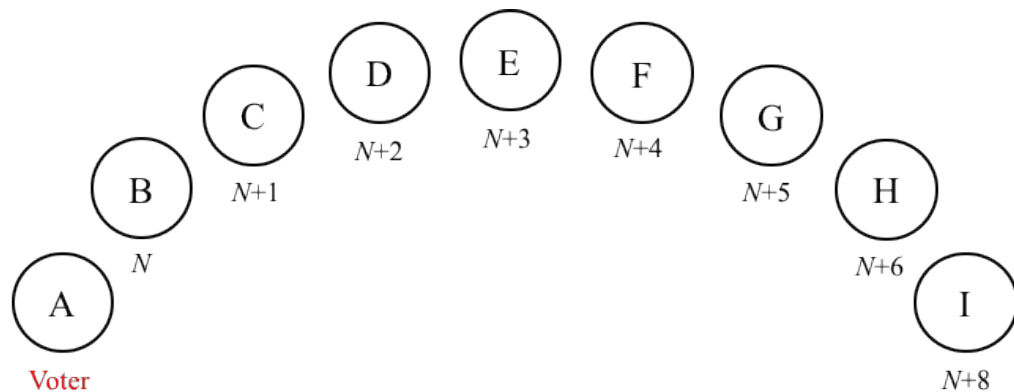
*The Money Tree for the Daytime Episodes*

<u>Daytime episode</u>
£1,000
£800
£600
£450
£300
£100
£50
£20

**Note.** This observation analysed the episodes from the Daytime episodes only. There were Champion League and Primetime episodes with different money trees allocation.

**Procedure.** The first-round of the WL TV game-show was observed. The voting behaviour of the contestants was recorded using an Excel spreadsheet. Each contestant's

position (A-I) and their gender were noted. The contestant who nominated another contestant was known as a *voter*. For the purposes of this thesis the recipient of the vote was referred to as the *votee*. The *voter's* vote for a *votee* was coded as an integer: its absolute number denoted the distance between *voter* and *votee* that ranged from one space away (direct-spatial neighbour) to eight spaces away ( $N, N+1, N+2, N+3, \dots, N+7$ ), its sign indicated the direction (+ve towards the right) or negative (-ve towards the left). To illustrate the data collection procedure, the contestants' positions were denoted alphabetically from A to I (i.e., nine contestants). As an example, Figure 3 shows contestant A as the assumed *voter*. The positions of other contestants were identified based on their distance from the *voter*, which defined the *voter-votee* spatial relationships. If contestant A voted contestant E as the weakest link, the *voter* had voted for contestant that was four spaces away ( $N+3$ ) from the *voter* and was coded as +4. Had contestant E voted for contestant A, then the vote was coded as -4.



*Figure 3.* The other contestants' position based on contestant A as the *voter*. This illustration is based on the WL game-show format. The contestants are standing in a semi-circle facing the game host at the centre.

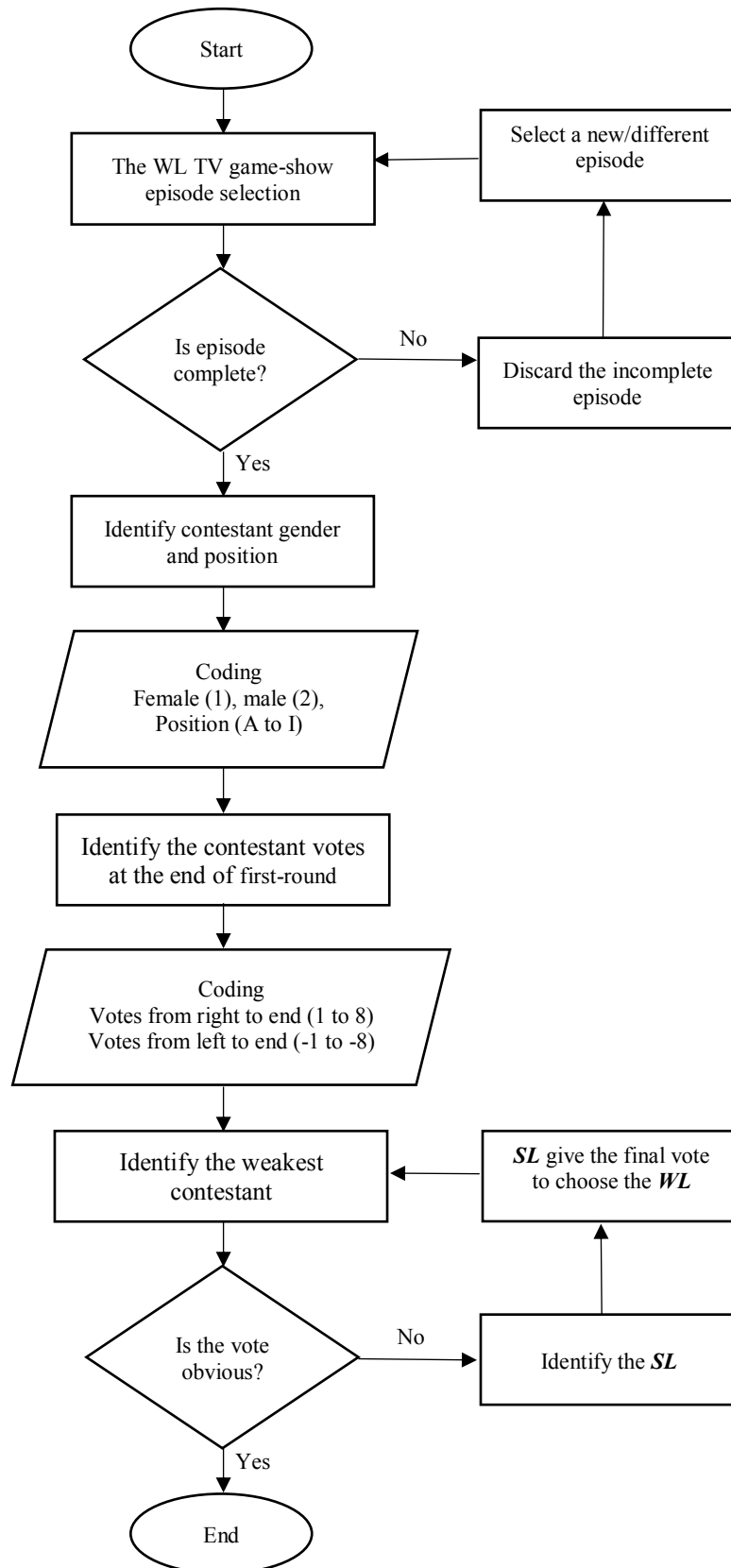


Figure 4. The observation protocol for the Weakest Link TV game-show.



There were two main systems used for coding episodes of WL. In the first 72-episode study, episodes were coded by independent raters that were undergraduate students using the episodes for their project work. The simplicity of the coding system, outlined above, made it easy to crosscheck their coding to ensure the accuracy of the coding procedure.

Figure 4 shows the observation protocol used for coding the 151-episode study. The observation protocol is shown in a flow chart. The protocol was used as a process to collect the raw data from the recorded videos.

### M1.3 Pilot Studies

A range of methods were piloted as a way to replicate the WL condition. The pilot studies were tested, but not developed due to the impracticality of the format and the inconsistent data that they generated.

**Pilot laboratory studies.** Small groups of students (5 to 10 participants) were recruited with an attempt to replicate the WL condition. One method involved carrying out a variant of the first-round of the WL game-show that fell short of actually labelling a participant as the *weakest link*. In order to get data that were relevant statistically, several rounds were run using the same participants. However, it soon emerged that participants managed their voting by choosing a different peer on each round. Therefore, voting performance was based on participants switching their vote from one peer to another and were not based either on the performance of other participants or their spatial positions. These studies suffered too because often participants were already known to each other, they were time consuming and they lacked the ecological validity of the actual WL game-show with its payoffs and studio atmosphere.

Similar biases emerged when different game scenarios were adopted (e.g., popular parlour games or children's party games like "Wink Murder", and variations of the WL with a participant playing against pre-set fictitious contestants).

The groups were required to *pick* someone within the group and later were asked to give reasons for their decision. The results derived from the pilot investigations were

inconsistent as the practicality of the study were not be able to be replicated on a larger scale. In comparison to the field experiments that made use of first-round voting in WL game-show, data was recorded for a large sample of 2,007 participants. The smaller group pilot studies were unable to generate the same volume and consistency of observations.

**Pilot Field experiment: - Fifteen to One TV game-show.** Ten episodes of the Fifteen to One (15-to-1) game-show were observed. The game-show layout was similar to the WL game-show, however, it had three rounds of questions and 15 contestants.

***First-round.*** In the first-round of questions each contestant had to field two questions each. An incorrect answer on the first question resulted in the loss of a “life” (each contestant began with three “lives”) a second incorrect answer resulted in the loss of the remaining two “lives” and the contestant was eliminated.

***Second-round.*** In the second-round a contestant was asked a general knowledge questions and any incorrect answer caused them to lose one “life”, and as before once a contestant had lost three “lives” they were eliminated. The round finished when there were only three remaining contestants. The feature of interest occurred in this round when a contestant had answered a question correctly they could nominate which of their peers should field the next question. The nominations made by contestants were measurable in the same way that contestants’ votes in the WL game-show were. Similar predictions were made regarding the NE. Nominating a peer to receive a question was construed as a negative act and therefore nominations for neighbours would be expected to be significantly less frequent than for non-neighbours.

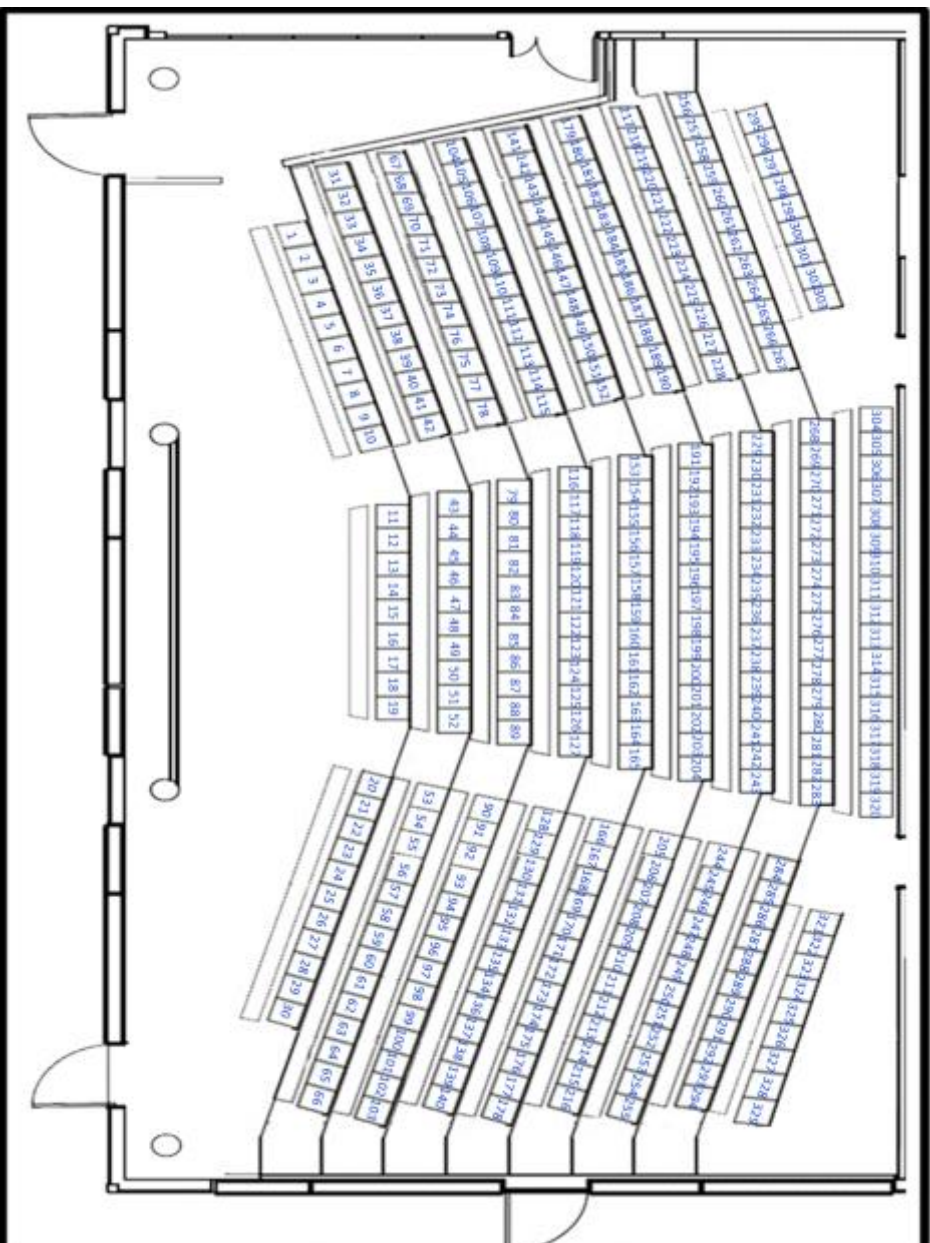
However, there were several limitations coding and interpreting the outcomes of the nominations in the 15-to-1. Firstly, the outcomes of the first-round in 15-to-1 were not stable in the same way that voting in the first-round of WL was stable. In every episode of the WL observed, the first-round resulted in each contestant making a single vote and a single contestant eliminated. In 15-to-1, the outcome differed from one episode to the next. This made models of expected frequencies in WL relevant to all episodes but in 15-to-1 the models were unique to each episode. In the second-round of questions in 15-to-1, nominations were counted as a function of the relative spatial positions of the contestants, but as contestants were continually eliminated during the round, the respective neighbour relations changed from one question to the next question. This means that the number of contestants that participated in the second-round also varied. This is one of the major differences in comparison to the WL game-show. The derivation of the expected voting pattern in 15-to-1 was convoluted, laborious and ambiguous.

Another confounding factor in 15-to-1 was the nomination for a contestant in 15-to-1 was not to eliminate them from the game. Thus, the nominated contestant was able to answer their question correctly, and hence re-nominate the previous contestant (Tit-for-Tat) to answer the following question. These cyclical nomination patterns were noted in the game-show and in one series a rule change was implemented to remove this bias. This strategic nomination by contestants served as an unregulated bias in interpreting the spatial nomination patterns in 15-to-1, hence, it meant that the nomination was no longer due to spatial position bias, but a strategic decision. This analysis was not taken further.

## M1.4 Lecture based Field Experiments

There were three field experiments: two of them were variants on a *voting activity*, the other used a variety of PD. This section describes the field experiment setting and set-up that were common for all the field experiments and then the voting activity and PD experiments are considered in turn for three different student cohorts. The final part considers the ethical considerations. The specific demographics of the participant samples used are detailed in the appropriate Results chapters (refer to **R2** and **R3**).

**Field experiment set-up.** The field experiments were conducted in a lecture theatre at the University of Lincoln with a capacity of 330 seats (refer to Figure 5). The lecture theatre comprised three blocks and seven rows, with each seat numbered row to row from 1 to 330. The seat numbers were placed at the back of the seat which were visible for participants. The lecture theatre was an ideal location for the activities, as it served as the location for the first introduction lecture for the first-year Psychology students. The putative purpose of the introductory lecture was to explain the School of Psychology tutorial system and *Credit Point* system. A scheme used to reward undergraduates with ‘credit’ for participating in studies conducted by staff and students in the School. Prior to the start of the activities, the participants were randomly given unique seat numbers, which required them to sit at the allocated seat as per the instruction sheet. The allocation was done by randomizing the seat numbers prior to the participants’ admittance to the lecture theatre. A team of ushers was used to hand out the seat allocation numbers to the students arriving for the lecture and to guide them to their unique seat. Once seated the respective activities took place.



*Figure 5.* The lecture theatre seating arrangement in the University of Lincoln. All 300 seats were numbered row by row. The seating layout was given to the participants in the instruction sheet in addition of the response sheet.

**Voting Activity Procedure.** The first field experiment was designed in the form of a *voting* task. The voting activity was set-up to get the participants to act as *voters* by nominating another participant as its recipient, the *votee*. The effect of the vote for the *votee* was to either increase or decrease their chances to win a subsequent lottery.

The number of lottery tickets participants had for a lottery for a prize draw was based on the votes they received in the voting activity. The rules of the *voting* task required each participant to cast either a positive (i.e., reward), neutral, or, negative (i.e., punishment) vote towards another participant. The vote choice was made in private by each participant who acted as a *voter*. A positive vote added to the chances of the *votee* (i.e., its recipient) winning a lottery for a prize draw. A negative vote reduced the chances of the *votee* winning the raffle. Neutral vote had no effect on the *votee*'s chance of winning. All participants were guaranteed an absolute minimum of at least one lottery ticket for the prize draw.

Table 2 shows the weight of the votes used in this activity.

Table 2

*The Type of Votes Performed by the Voter*

Vote	Value
Positive	+1 and +5
Neutral	0
Negative	-1 and -5

**Note.** The participants were instructed to vote either +1, +5, 0, -1 or -5. A positive vote of +5 increased the number of lottery tickets a *votee* received. A negative vote of -1 decreased their number of lottery tickets by one. Neutral votes had no effect on a *votee*'s stack of tickets. Each participant was guaranteed at least one ticket for the lottery. Tickets were virtual.

For the *voting* activity, each participant was required make three votes for another participant: (a) sitting in the same block and in the same row of seats, (b) sitting in the same block, (c) sitting anywhere in the lecture theatre. Participants also had an open-ended question that asked them to give a reason for their voting choice. However, the last task was not compulsory.

The participants were given a sealed envelope with two forms; instruction and response sheets. The instruction sheet explained the rules of the *voting* activity and a note for giving their consent with a right to withdraw from participating. Each participant from 2013/2014 and 2015/2016 cohorts received the same instructions but in a different order. Figure 6 shows the instruction given to the participants.

The response sheet served as their *ballot* paper (refer to Appendix A1-A10 for 2013/2014 cohort and B1 to B4 for 2015/2016 cohort), which included the seating plan (refer to Figure 5) for the lecture theatre. In the *ballot* paper, the participants were asked to identify their seat number (compulsory), age and gender. The participants were required to complete the voting tasks above (i.e., a, b and c). The value and valence of their vote were random (i.e. -5, -1, 0, +1 and +5). Votes were cast by marking an *X* on the seating plan and later the response sheet was placed back in the envelope, sealed and given to the researchers.

A simple database in Excel was used to code votes made. The content includes the lecture theatre layout with seat number and the measured variables. The data was later cross-checked to ensure the collected data was correct with no error.



**Procedure:** - We will ask you to make **THREE** votes that vary according to **spatial scale**.

I.) Vote for someone in the **LECTURE THEATRE** by making a cross on the seating plan.

II.) Vote for someone in your **BLOCK** by making a cross.

III.) Vote for someone in your **BLOCK** and in your **ROW** by making a cross.

VI.) MOST IMPORTANTLY, please indicate your seat number here and make a note of it:-

I AM SAT IN BLOCK: -

I AM SAT IN ROW: -

I AM SAT IN SEAT: -

### **RULES**

**Please be quiet until everyone has made their votes and the experimenter**

**The only rule is that you can't vote for yourself!!!!**

**Procedure:** - We will ask you to make **THREE** votes that vary according to **spatial scale**.

I.) Vote for someone in the **BLOCK** and in your **ROW** by making a cross on the seating plan.

II.) Vote for someone in your **BLOCK** by making a cross.

III.) Vote for someone in your **LECTURE THEATRE** by making a cross.

VI.) MOST IMPORTANTLY, please indicate your seat number here and make a note of it:-

I AM SAT IN BLOCK: -

I AM SAT IN ROW: -

I AM SAT IN SEAT: -

### **RULES**

**Please be quiet until everyone has made their votes and the experimenter**

**The only rule is that you can't vote for yourself!!!!**

*Figures 6.1 and 6.2.* The instruction given to the participants on how to take part in the voting task for 2013/2014 and 2015/2016 cohorts (see Appendix A and B). Figure 6.1 (upper panel) is the instruction given to the 2013/2014 cohort and Figure 6.2 (lower panel) is given to the 2015/2016 cohort. The differences were at the item I and III where the sequences of the voting task were switched.

**Prisoner's Dilemma Game Procedure.** This field experiment was conducted in a form of PD game with 2014/2015 cohort. The aim of this activity was designed to get the participants to “defect” or “cooperate” with another participant by playing either *tails* or *heads*, in a task called *Coin's game* (CG).

The rules in the CG required the participants to play against another participant in the lecture seated: (a) the same block of the lecture theatre and in the same row of seats, (b) the same block of the lecture theatre and one row at the front or back, and (c) anywhere within the lecture theatre.

The activity was performed individually and in private. The task required them to choose either *Heads* (H) or *tails* (T) as per coin flipping task. In this activity, the participants were presented with four different payoffs as showed in Figure 7, the payoffs matrix for the CG.

		Player X	
		<i>heads</i> (H)	<i>tails</i> (T)
Player Y	<i>heads</i> (H)	(3, 3)	(0, 5)
	<i>tails</i> (T)	(5, 0)	(1, 1)

*Figure 7.* The payoff matrix for the coin's game. The first number in the brackets gives player *Y*'s payoff and the second number gives player *X*'s.

In the CG activity, the participant played as player *Y* and were asked to maximize their payoff by playing with another participant as player *X*. In this context, player *X* (another participant) was assumed to be seated either at; (a) the same block of the lecture

theatre and in the same row of seats, (b) the same block of the lecture theatre and one row at the front and back or (c) within the lecture theatre.

Each participant was given a sealed envelope with a form (response sheet) that explained the CG activity. The response sheet contained two sections; (a) the activity instruction and the consent for a right to withdraw from the activity and the response sections. Following instruction was given to the participants (see Figure 8) prior to the start of the activity.

**We are going to ask you to play the ‘coin game’ 3 times.**

**Each time you play, you will indicate how your play by circling ‘H’ or ‘T’ or entering it into a text box, on the other side of this ‘Response Sheet’. Your responses are secret and anonymous and other players will not see what you have done. You will never get to see how other players responded.**

**Your aim is to try and score as many points as possible. We will work this out by adding up how many points you scored in the 3 games and dividing it by how many players there were that you played against. This will give us an average point score for each participant.**

**When you have made your responses we will give you a verbal debrief about what the study is about, but in short there are no catches or tricks involved. Computer scientists have generated computer programs that have played these kinds of games before. We want to know if a group of people play the game the same way as the computer program.**

*Figure 8.* The instruction given to the participants on how to take part in the coin’s game for 2014/2015 cohort (see Appendix C).

Whilst, in the response section (refer to Appendices C1 to C6), the participants were asked to identify their seat number (compulsory), age and gender. It also enclosed the information on the CG activity where each individual participant was required to

complete a form with six compulsory questions and one optional open-ended question which related to the payoffs matrix in Figure 7.

The first question centered on their self-assessment on how well they thought they had understood the rules of CG (i.e., Fairly well, well, OK, Not very well, Not at all). The participant required only to answer one which best fit their current understanding after having been briefed. Two screening questions were asked next. Participants had to respond by referring to the payoff matrix, if they were to play against player  $X$ . The questions were asked in a condition where the player  $X$  played either H or T. The participants were required to choose only one option, in which, between H and T as to their best response. The screening questions were another check that participants understood the CG task.

***The game.*** Two questions were described in six different sequences (refer to Appendices C1-C6 for different versions). The versions were randomly assigned to participants where the versions were varied to test participants' decision based on (a) playing against participants at lateral (left/right) position, (b) playing against participants at longitudinal (front/back) position and (c) playing against participants at lateral and longitudinal position. Each of the question asked the participants to make a choice either to play H or T with direct-neighbour(s) and indirect -neighbour(s) two-spaces away. The last question was a standard question across the versions to ask the participants to choose playing H and T against their direct-lateral neighbour(s) or somebody else within the lecture theatre. An optional last open-ended question invited participants to give the reasons for their choices. The sealed envelope contained the instruction and response sheet

were distributed randomly among participants. Upon completion, the response sheet was placed back in the envelope, sealed and given to the researchers.

**Ethical Consideration.** Ethical approval was received from the School of Psychology Research Ethics Committee to conduct the studies. The participants were given consent forms and prior to the activities, they were briefed and debriefed after the session end. Participants were told that they were free not to complete the exercise and if this could be done simply by not filling out the form. The participants were advised that they had a right to withdraw from participating in the studies. No personal details were recorded (e.g., name). All participants received a credit point for participating even if they chose not to complete the task.

## Chapter 3

### Results

The **Results** section comprises three main parts (**R1**, **R2** and **R3**) to reflect the different methodologies used to collect the quantitative data for this thesis. The first part is made up of three main studies presenting the findings from observational field studies used to analyse voting behaviour of contestants on the TV game-show the WL. The second and third part built on the findings from the observational field studies to generate hypotheses that were tested using two experiments carried out in the setting of a University lecture.

# Chapter R1:

## Demonstration and Replication of The Neighbour Effect in The Weakest Link TV Game-Show using Observational Field Studies

“In addition, this game-show (WL) provides an ideal laboratory to study human decision-making. The rules are well-defined and the stakes are high, something that is not easy to replicate elsewhere” (Barmish & Boston, 2009, p. 3).

### R1.1 Chapter Outline

The chapter is presented as three studies with the reanalysis of the data used in the original study by Goddard et al. (2011) being followed by a new replication. The final study analysed second-round voting.

The starting point for this research was an observation that contestants on a TV game-show showed a significant reluctance to nominate the contestants that were next to them as the worst contestant, the so called weakest link (Goddard et. al, 2011). This was referred to as the *neighbour effect* (NE). The first study presents a 72-episode study where the original data was reanalysed. It confirmed the original NE but went further by also showing a significant gender effect in the first-round voting behaviour not originally noted by Goddard et al.

To test the NE and the unexpected gender effect, the second study used a 151-episode replication by analysing the first-round voting responses made by contestants in this different, extended set of WL episodes. It was found that the NE was still present but a different kind of gender effect was noted. This was interpreted as there being a robust NE but the gender effect that emerged in the original data set (72 episodes) changed using the larger set (151 episodes).

The third study analysed the second-round voting pattern from the original 72-episode study. This confirmed that the proximity and NE carried over from the first-round to the second-round. Furthermore, it revealed another two biases in voting behaviour. The second-round *voters* were significantly more likely to retaliate against the contestants that had voted against them in the first-round. Those second-round *voters* that had picked a *votee* in the first-round appeared to be significantly more likely to vote against them again in the second-round voting.

#### R1.1 Chapter outline

#### R1.2 Observational Field Study 1: Original Demonstration of the Neighbour Effect

#### R1.3 Observational Field Study 2: Replication of The Neighbour Effect

#### R1.4 Observational Field Study 3: Analysis of Second-round Voting Reciprocity,

#### Retaliation and Consistency

#### R1.5 General Discussion



## R1.2 Observational Field Study 1: Original Demonstration of the Neighbour Effect

**Abstract.** The pattern of voting in the first-round of 72 episodes of the UK version of the TV game-show The Weakest Link (WL) was analysed. In each episode, the first-round finished when each of the nine contestants had picked one of their peers as the WL. The observations were based on the original data set used by Goddard et al. (2011). Four significant main findings emerged; (i) contestants avoided voting their direct-neighbour (s) as the WL, (ii) however, *voters'* likelihood to vote for a peer was not a simple linear function of distance per se, (iii.) women were more likely to be voted as weakest link (iv) the “neighbour-avoidance” effect increased as the group consensus as to the identity of the weakest link decreased. These findings replicated the original conclusions of Goddard et al. but extended their observations by demonstrating a significant gender effect in the voting patterns. The gender effect was opposite that found in similar studies using the US version of WL. This suggests that the pattern of voting is more complex than originally thought.

**Introduction.** The configuration of contestants in the WL makes it easy to define the spatial relationships between all contestants. Furthermore, each contestant made a discrete decision at the end of the first-round of questions by voting for one of their peers. Therefore, the design of the game-show conveniently affords a means to examine whether voting behaviour was affected by the contestants' spatial arrangement.

The null hypothesis was defined by the voting levels expected purely due to chance. For example, in a show with nine contestants (i.e., labelled A-I and arranged according to Figure 3) the probability that contestant A vote for any of the other contestants is  $\frac{1}{8}$  or .125. With respect to the voting pattern predicted for neighbours, contestants A and G only have one neighbour each (B and H respectively), so  $p(\text{neighbour}) = .125$  for each of them. Contestants B through H, each have two neighbours each so  $p(\text{neighbour}) = 2 * .125$  for each of them. Therefore, in a single episode the expected number of votes for the neighbour is two, simply the sum of the individual probabilities for each contestant voting their neighbour ( $E(\text{neighbour}) = (2 * .125) + (7 * .250) = 2$ ). Across the 72 episodes observed, the total expected votes for the direct-neighbours based solely on simple probability was  $2 * 72 = 144$ . The same reasoning was applied to each of the other possible *voter-votee* spatial relationships to derive expected vote frequencies.

It follows that if the contestants were influenced by the spatial proximity of their fellows then voting frequency should vary as some function of *voter-votee* separation. One *prima facie* prediction was that contestants would be less likely to vote for their close neighbours so more likely to vote for those further away, giving an inverse relationship between vote frequency and proximity. This neat linear relationship was originally predicted by Goddard et al. (2011), the gradient of which would provide a measure of the strength of the proximity effect. It turned out as it happened that there was a strong effect of *voter-votee* proximity on vote frequency. Most notably, *voters* were averse to nominating their direct-neighbours as the weakest link. However, the contestants furthest away received the number of votes expected purely due to probability. Those contestants

between these extremes accrued more votes than would be expected. Therefore, overall, there was a significant departure between the observed and expected voting patterns. It is described here in terms of three phases *proximal-medial-distal*. Goddard et al described it as a neighbour-avoidance effect, NE, (later this becomes a NE because of observations regarding valence) that because of their proximity *voters* avoid doing some bad to their direct-neighbours. They offered no explanation for the *medial/distal* effects.

The method used in the research presented in this chapter was an observational field study of the UK version of the WL game-show. The first part reanalyses and extends Goddard et al's (2011) original study designed to test the relationship between vote frequency and the physical proximity between *voter* and *votee*. These observations were extended by including an analysis of the gender of the participants to examine whether gender effects similar to the ones exposed by Levitt (2004) and Antonovics et al. (2005) would also emerge in the UK version of the game-show. The study used the same original coding of the data provided by Goddard et al. One of the main aims of this opening study was to provide the researcher with an opportunity to scrutinise an example of previously coded data as a preliminary to carrying out a more comprehensive replication using a new set of episodes of the WL (i.e., the same method was used to code and analyse the voting patterns of contestants in 151-episodes of the UK version of WL, described in **M1.2**).

Thus, the objectives of this study were:

1. To investigate the relationship between the negative voting pattern and proximity.
2. To understand the consensus (i.e., endogenous factor) effect towards voting outcome. A key factor that lead Goddard et al. (2011) to attribute a NE was responsible for the voting bias they found, was because the effect was related to the consensus regarding the identity of the WL. This relationship should become more pronounced as the consensus regarding the identity of the WL decreased.
3. To explore gender (i.e., exogenous factor) effect in voting behaviour.

**Method.** The methods used here have already been described in the General Methods, chapter **M1.2**. This section provides a summary of relevant details pertaining to a reanalysis of Goddard et al. (2011).

***Field observation of WL TV game-show.*** The first-round voting patterns in 72 episodes of the UK version of the WL game-show were analysed. The episodes were originally shown on free to air TV in the UK TV between 24<sup>th</sup> April 2002 and 8<sup>th</sup> January 2003. Each episode was coded according to the date it was aired using a team of independent coders who were undergraduate students at the University of Lincoln. The coders undertook the task as part of project work towards year III dissertations. Few errors were identified in the coding but any discrepancies that did arise were easily corrected at the time by recourse to the original episodes. This part of the study was based on the

coding originally undertaken by the independent coders and was not based on observing the original 72 episodes of the WL.

This preliminary study, used 72 episodes of the WL, each of which included nine contestants in each episode, hence a total of 648 (Female = 300, Male = 348) contestants voting activities were recorded altogether. This study used first-round WL TV game-show voting, where the contestants each made an eight-alternative-forced-choice (8-AFC) to vote out another contestant as WL. Each contestant's position (A - I, see schematic at Figure 3) and their gender were noted, as well as their vote. The votes were recorded as integers such that their absolute value represented the distance between *voter-votee*, from one space away (direct-spatial neighbour) to eight spaces away (  $N, N+1, N+2, N+3, \dots, N+7$  ). The number's sign represented the direction from the *voter* to the *votee* (+ve towards the right or -ve towards the left). For example, the *voter* in position A, nominating *votee* in position I, eight spaces away, was coded, 8. Similarly, voter in position I, picking *votee* in position H was coded -1.

Other demographic factors (e.g., age, marital status, ethnicity, education, income and occupation) were not recorded. Other game-specific information was also not recorded such as number of questions answered, their accuracy, banking decisions, amount won per episode were not recorded.

**Results.** The analysis used in this section always followed the same pattern in this replication of Goddard et al. (2011). The *observed* frequency of votes actually made by contestants was counted and compared with the frequency of votes *expected* purely due to

probability (Goddard et al.). This section is structured so that the first part explains how the *expected* frequencies were derived. The second and third parts replicated the analysis of Goddard et al. regarding voting behaviour and proximity followed by proximity and consensus; the third part extended the observations of Goddard et al. by testing for a gender effect.

***Deriving expected frequencies.*** Expected voting frequencies were derived by recourse to rudimentary probability theory. By assuming a random unbiased model then the probability associated with a contestant voting for any of the other eight contestants was always  $\frac{1}{8}$  (.125). As the contestants at the endpoints of the array, A and I were not considered as spatial neighbours (even though they are strictly temporal neighbours in the sequence of questions) meant that they only had one direct-neighbour each, whereas contestants B through H had two each. Therefore, the expected frequency for contestants voting for their neighbour in a single round was two, simply the sum of the probabilities of individual contestants voting for their neighbour ( $2 \times .125 + 7 \times .250$ ). Expected frequencies were calculated for all spatial relationships, neighbour, neighbour +1, ..., neighbour + 7, giving, 2, 1.750, 1.500, 1.250, 1, .750, .500, .250 respectively. Notice that these sum to nine, the same as the number of contestant votes in a single round. Over the course of 72 episodes the expected number of neighbour votes was 144.

Expected frequencies for other *voter-votee* relationships ( $N, N+1, N+2, N+3, \dots, N+7$ ) are shown below in Table 3, calculated using the same procedure to derive the expected frequency. Notice that commonest contestant spatial relationship is the *neighbour* ( $N$ ), with 16 neighbours forming 8 pairs. The next-door-neighbour-but-one

( $N+1$ ) is the next commonest with 7 pairs, dropping successively, to the least common single pairing between the most distant contestants, A and I, eight spaces away ( $N+7$ ) from each other ( $N = 8$ ;  $N+1 = 7$ ;  $N+2 = 6$ ;  $N+3 = 5$ ;  $N+4 = 4$ ;  $N+5 = 3$ ;  $N+6 = 2$ ;  $N+7 = 1$ ).

*Expected frequencies with weakest link and without weakest link.* Expected frequencies were calculated both including and excluding the vote of the eventual weakest link in the first-round.

Table 3

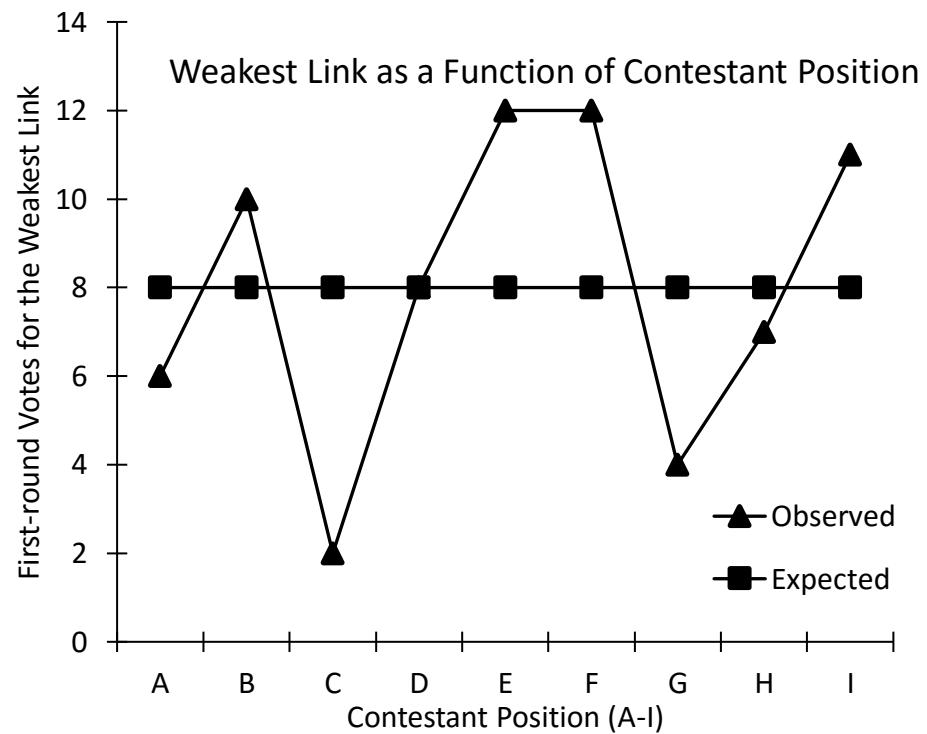
*The Calculated Expected Frequencies for Each Voter-Votee Position*

Pos.	A	B	C	D	E	F	G	H	I	Freq.
$N$	0.125	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.125	2.000
$N+1$	0.125	0.125	0.250	0.250	0.250	0.250	0.250	0.125	0.125	1.750
$N+2$	0.125	0.125	0.125	0.250	0.250	0.250	0.125	0.125	0.125	1.500
$N+3$	0.125	0.125	0.125	0.125	0.250	0.125	0.125	0.125	0.125	1.250
$N+4$	0.125	0.125	0.125	0.125	-	0.125	0.125	0.125	0.125	1.000
$N+5$	0.125	0.125	0.125	-	-	-	0.125	0.125	0.125	0.750
$N+6$	0.125	0.125	-	-	-	-	-	0.125	0.125	0.500
$N+7$	0.125	-	-	-	-	-	-	-	0.125	0.250

**Note.** Position (Pos.) for each contestant was denoted by A to I and contestants' votes towards other contestants represented with  $N$ ,  $N+1$ ,  $N+2$ , ...,  $N+7$ . Each frequency (Freq.) was calculated based on the rudimentary probability theory.

**Weakest link as a function of contestant position.** One assumption is made that there was no systematic bias in the arrangement of the contestants in the game-show. To test this the position of the WL throughout the 72 episodes was recorded and the frequencies that first-round weakest links appeared in positions A - I is shown in Figure 9. Over the course of 72 episodes and assuming that, there was no significant bias in the positioning of the weakest link then there would be no predicted significant deviation from

the expected eight occurrences of the weakest link in each position  $\frac{72}{9} = 8$ , square symbols). There was no systematic bias in the spatial or temporal order of candidates as the observed frequencies ( $f_o$ ) of the weakest link found in each spatial position (A-I) compared to the expected frequency ( $f_e$ ) of 8 ( $n(\text{episodes}) / n(\text{positions})$ ;  $\frac{72}{9} = 8$ ) was non-significant ( $\chi^2(8, N = 72) = 12.750, p = .121$ , small effect size, Cramer's  $V = .149$ ). Contestants were no more or less likely to vote for candidates that preceded or followed them temporally in the testing sequence.



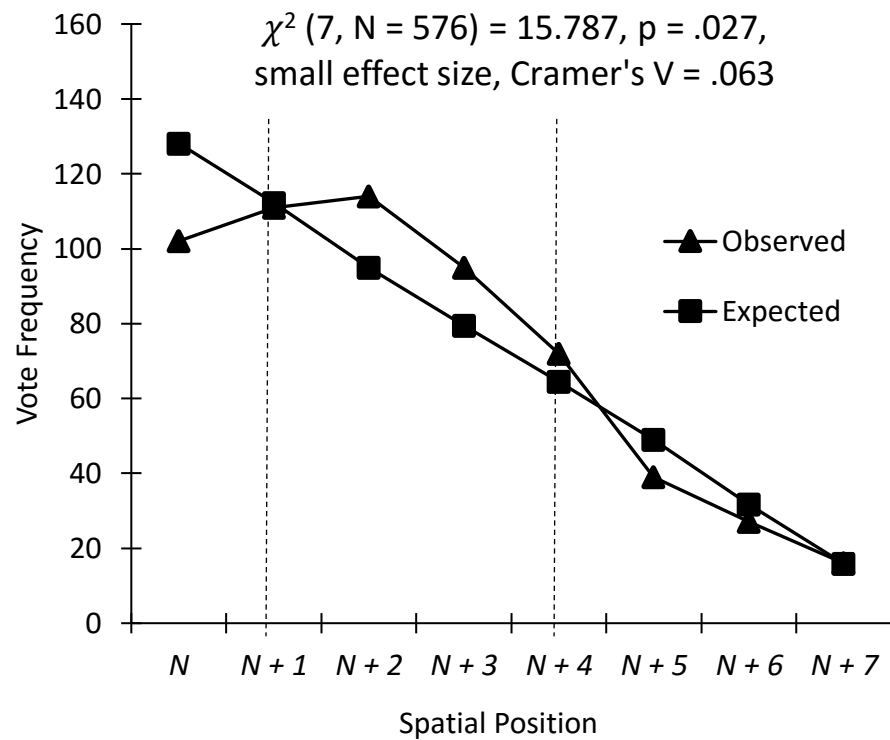
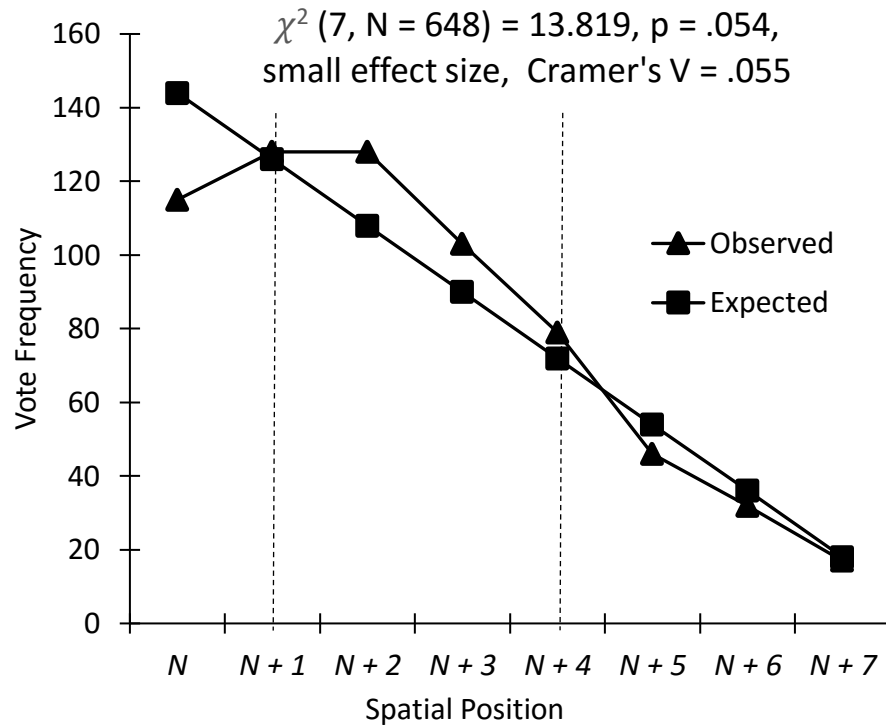
*Figure 9.* Plots the frequency that the weakest link landed in each of the nine contestant positions (A - I). The square symbols show the expected distribution and the triangle show the observed frequencies.



***Finding 1: Voter-votee proximity (The “neighbour-avoidance” effect).*** The main finding of Goddard et al. (2011) was an effect of proximity on voting behaviour.

Figure 10.1 (upper panel) shows the expected pattern of voting expected (square) and observed (triangle) for the contestants following the first-round of questions (The data shown here includes the votes cast by the eventual weakest link). The expected frequencies were derived using the method described earlier and the observed frequency was a simple count of votes for the specified *voter-votee* relationships. The key features of the figure show that vote frequency seems to be described by three distinctive phases. The *proximal*/neighbour phase shows that direct-neighbours avoided voting for each other ( $N$ ). The *medial* phase ( $N+1$  to  $N+4$ ) shows where contestants were most vulnerable to being voted whilst those furthest away in the *distal*/distant phase ( $N+5$  to  $N+7$ ) showed the expected vote frequencies.

The voting frequencies including the weakest link votes change the voting pattern slightly such that the spatial effect is non-significant but the neighbour *versus* non-neighbour remains significant. If implicit processes operated to influence decision-making, then they should become apparent as systematic biases over the course of many decisions. In the context of the weakest link, decisions were the votes cast measured as a function of the spatial relationship between the voter and the candidate.



Figures 10.1 and 10.2. shows the voting frequency as a function of spatial position for 72-epside study. It compares the expected voting frequency (square), derived as the unbiased pattern of voting, with the observed pattern of voting (triangle). Voting frequency is

plotted as a function of the spatial relationship between *voter* and candidate where ' $N$ ' refers to the *voter*'s neighbour and ' $N + 1$ ' the next-door-but-one neighbour and so on. The upper panel (Figure 10.1) is the voting pattern which includes the votes cast by the eventual weakest link while the lower panel (Figure 10.2) is the voting pattern which excludes the votes cast by the eventual weakest link.

Significant departures between the unbiased pattern of voting and the observed pattern of voting can be seen in Figures 10.1 and 10.2 where contestants were less likely to vote for their direct-neighbour,  $N$ . Although this supports the predicted avoidance of voting for neighbours, Figure 10.1 also shows that there was not a simple linear relationship between voting pattern and proximity as more distant candidates were not necessarily more likely to acquire votes (The data shown here includes the votes cast by the eventual weakest link). A comparison between voters' 'neighbour votes' against 'non-neighbour votes' was significant ( $\chi^2(1) = 7.508$ ,  $p < .005$ ).

**Finding 2: The effect of gender.** A simple analysis across the 72 episodes reveals that the gender of the weakest link in the first-round voting was a woman on 46 occasions and a man in only 26 episodes. This is surprising because there were more male's contestants than females. If this followed the composition of the sample then men would be the weakest link in 38.667 episodes with 33.333 female weakest links ( $\chi^2(1, N = 72) = 8.993, p < .005$ , medium effect size Cramer's  $V = .353$ ). On first sight, it seems that at least for the first-round of voting there is marked discrimination in voting the weakest link based on gender. Breaking these results down further, Figures 11, 12.1, 12.2, 13.1 and 13.2 show the observed and expected vote frequencies for female and male voters voting for females and males respectively.

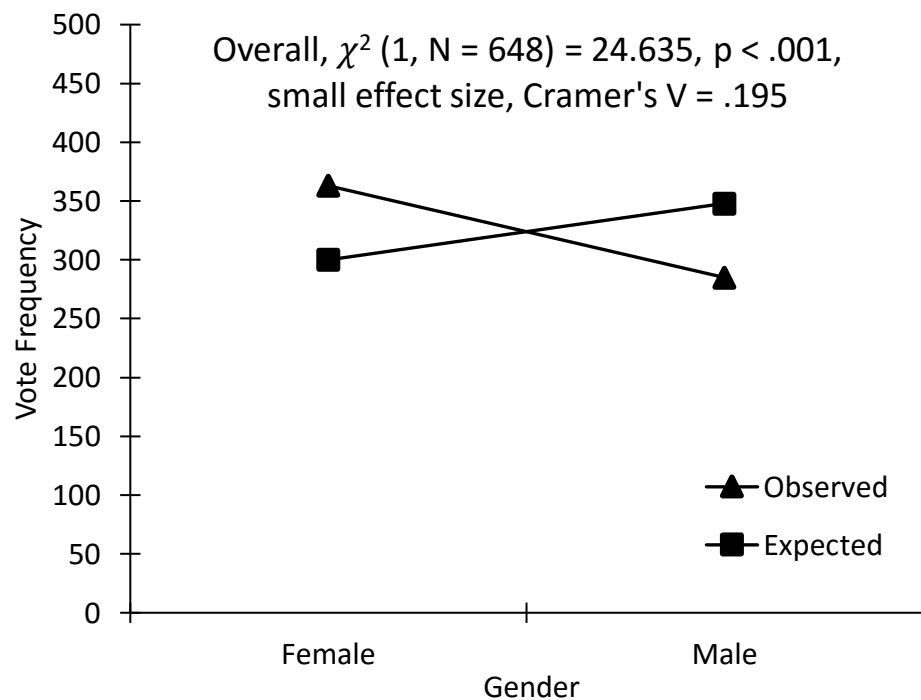
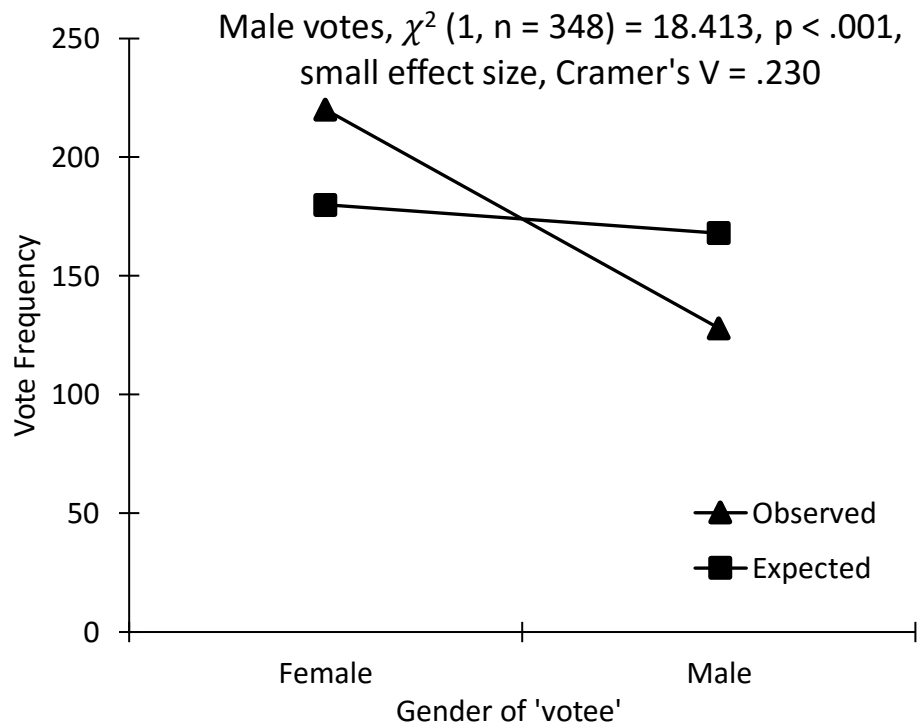
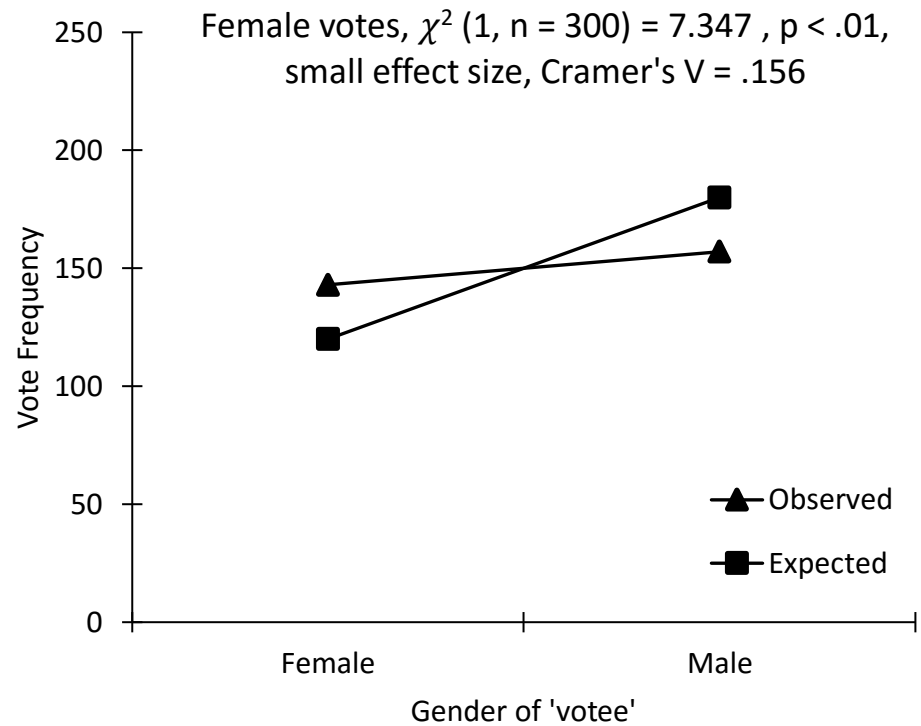


Figure 11. The vote frequencies with respect to female and male contestants for 72-episode study. The triangular is the observed vote frequencies and the square is the expected vote frequencies by each gender.



*Figures 12.1 and 12.2.* Voting performance for both gender in 72-episode study. Figure 12.1 (upper panel) is a voting performance for female contestants and Figure 12.2 (lower panel) is a voting performance for male contestants. The triangular is the observed vote frequencies and the square is the expected vote frequencies.

Figure 11 shows that females in the first-round of voting were subject to significant negative bias and discrimination by both male and female *voters*. This result was a surprise. Firstly, the predicted bias was in terms of *voter-votee* spatial relationships rather than gender. Secondly, although gender effects had been noted previously they were in the opposite direction. Particularly in the case of women *voters* showing a preference-based discrimination for women and against men. Antonovics et al. (2005) used the Group Bias Statistic (GBS) to measure group bias statistic and found that for women voting for other women the GBS in first-round voting were .739 ( $n = 222$ ,  $sd = 1.41$ ) and .783 ( $n = 111$ ,  $sd = 1.086$ ) for the different types of WL show. Applying their metric to Goddard et al.'s (2011) data gives a GBS of 1.21 ( $n = 300$ ) in the opposite direction (recall that 1 = no discrimination). The GBS statistic for males GBS were non-significant revealing no gender based discrimination in the U.S versions of WL, .976 ( $n = 222$ ,  $sd = 1.22$ ) and 1.00 ( $n = 113$ ,  $sd = 1.185$ ), compared with the GBS of 0.764 ( $n = 348$ ) indicating strong out-group discrimination in the UK version of WL. Levitt (2004) found "...women receive weakly fewer votes than men..., although in no case is the difference across gender statistically significant" (Levitt, 2004, p. 446).

The possible source of the discrimination found in Goddard et al. (2011) data is important because it introduces the possibility that the significant differences that emerged were not simply attributed to *voter-votee* proximity, but represent a more complex picture of bias regarding proximity and gender affecting voting in which regarded as the endogenous and exogenous factors. They are also important as a potential way of investigating discrimination something overlooked in their original analysis.

*Distribution of men and women.* Aside from the advantages highlighted in **II.2** detailing the advantages of studying game-shows, a glaring disadvantage is that the researcher has no control over the selection of the contestants and their distribution in the setting of the game-show. Table 4 presents the arrangement of contestants by gender in the shows configuration, shows that the distribution is not random. Of the 72 episodes, 60 had five males and four females with the remainder having five females and four males.

Table 4

Contestant Arrangement by Gender (72 - episode Study)

<u>Arrangement</u>	<u>Number of episodes</u>
mfmfmfmfm	29
mfmfmfmffm	5
mfmffmfmfm	5
mfmfmffmm	6
mfmffmfmfm	3

*Note.* The remaining 24 episodes comprised 16 arrangements that occurred once and four that occurred twice.

Table 4 shows that the distribution of contestants was not random. The commonest pattern that accounted for 40% of the episodes interleaved male and female contestants (i.e., Male-Female-Male-Female-Male-Female-Male-Female-Male). The vast majority of the other combinations were variants of this form also, meaning that in the first-round the neighbours of the male contestants were females and vice versa. This raised the real possibility that what Goddard et al. (2011) interpreted as a proximity based voting bias might actually be the manifestation of a gender effect allied with the non-random distribution of contestants by the show's producers.

*Gender and proximity.* Both males and females demonstrated discrimination by significantly voting against women. Women make up the majority of the neighbours for male contestants and it follows that males make up the majority of neighbours for female contestants. If Goddard et al. (2011) NE is just gender bias, then male and female *voters* should show different voting patterns as a function of *voter-votee* spatial relationships. Female *voters* should demonstrate a strong NE because their neighbours are males and they have a voting bias against females, therefore they will be more likely to vote for females that are not near neighbours. Males on the other hand should show no NE, or if anything a reversed NE, because their neighbours are females and males are biased against them.

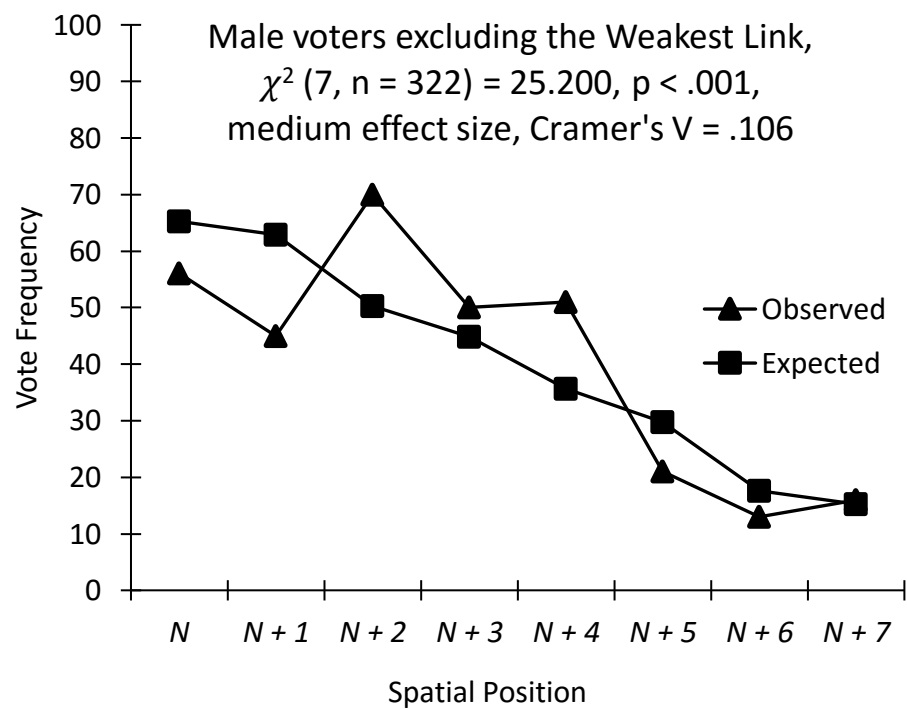
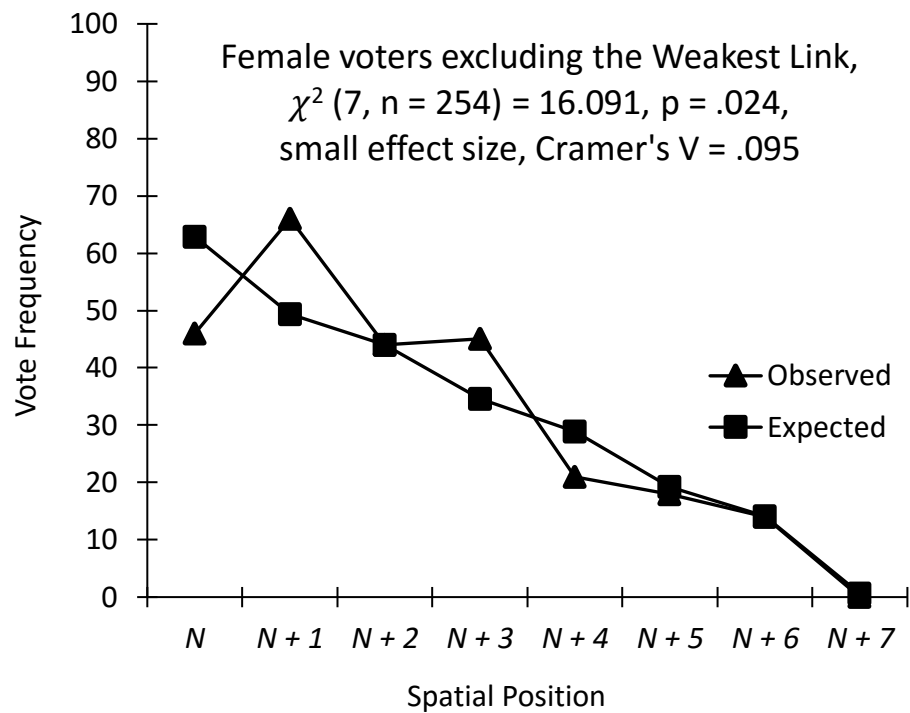
Figures 13.1 and 13.2 show these predictions were not completely borne out. Figure 13.1 (upper panel) shows the female voters' voting pattern as a function of *voter-votee* spatial relationships. A significant departure from the expected is noted and alternates about the expected performances according to the predominate gender of the *votee*. Where *votees* were predominantly male ( $N$ ,  $N+2$ ,  $N+4$ ,  $N+6$ ) *voters* elected them fewer times than would be expected. When the votes were predominantly female ( $N+1$ ,  $N+3$ ,  $N+5$ ,  $N+7$ ) *voters* tended to vote them as the weakest link more than expected. This causes the observed voting pattern to alternate above and below the expected pattern. This shows the expected pattern for a NE but can also be accounted for by the biased voting of the female contestants.

Figure 13.2 (lower panel) shows the performance of the male *voters* using the same axis. They also demonstrate that the expected pattern of performance was not a good fit



for the observed pattern of voting. Given that males also show a strong discrimination against women, and their direct-neighbours were predominantly females then it was a surprise that they voted for their female neighbours fewer times than would be expected. If anything, they would be expected to vote for their female neighbours more than would be expected. It seemed that males still demonstrated the NE. The gender effect is still visible by the alternating pattern of the observed voting aside from the neighbour position ( $N$ ). When the *votees* were predominantly male however ( $N+1, N+3, N+5, N+7$ ) *voters* tended to vote for them as weakest link fewer times than expected, but when predominantly female ( $N+2, N+4, N+6$ ) they voted them weakest link more than expected.

This section suggests that the voting in WL is subject to at least two forms of bias regarding gender and proximity. The next section follows Goddard et al. (2011) analysis of the proximity effect as a function of the consensus of the other *voters* towards the WL.



Figures 13.1 and 13.2 show the votes by gender as a function of spatial position in 72-episode study. Figure 13.1 (upper panel) is the female voters and Figure 13.2 (lower panel) is the male voters. The triangular is the observed vote frequencies and the square is

the expected vote frequencies. Figures 13.1 and 13.2 show vote frequency as a function of *voter-votee* relationships for male and female voters. Notice in both cases the NE is apparent at  $N$ .

***Finding 3: The effect of consensus.*** Consensus refers to the number of votes accrued by the eventual weakest link. The assumption was that in a round where all eight contestants (i.e., aside from the weakest link themselves) voted for the same *votee* then there was some degree of certainty over the identity of the weakest link. Likewise, low consensus rounds were indicative of uncertainty. Proximity biases would be more likely to be found in circumstances of uncertainty therefore the NE should increase with a decrease in consensus. Table 5 below shows the distribution of consensus across the 72 episodes.

Table 5

*The Contestants Consensus Votes in 72 Episodes*

Consensus	Frequency
1	0
2	1
3	12
4	16
5	17
6	11
7	12
8	3
Total	72

***Note.*** Shows the frequency distribution of episodes according to consensus, the number of votes cast for the eventual weakest link during the first-round of voting.

*Generating the null hypothesis for the distribution of votes according to consensus.*

The format of the WL game-show had nine contestants with each selecting one other contestant as the *weakest link* from a menu of eight choices. Therefore, each contestant had a slightly different choice-set as a contestant was unable to vote for themselves as the weakest link. This gave a total number of unique voting permutations for the show of: -

$$8^9 = 134,217,728$$

The number of permutations that resulted in any one of the contestants, say contestant A, as being the “Weakest Link” was therefore: -

$$\frac{8^9}{9} = 14,913,080.9$$

The generation of the null hypothesis for the distribution of episodes as a function of the consensus for votes for the *weakest link* was calculated in two steps. In the first step, simple probability theory was used to establish the frequency of episodes that resulted in  $n$  votes for a particular contestant, say the contestant in position A. For example, when the consensus was the maximum of eight, such that all eight contestants were in agreement by voting for the *weakest link* as the contestant in position A, then there were eight ways in which this could possibly happen. The contestants in positions B through I must have all voted for A. The contestant in position A, unable to nominate themselves, voted for each of the other eight contestants in positions B through I in turn.

Therefore, there were eight permutations where it was possible for the contestant in position A to be voted as the *weakest link*. Extrapolating this for the other positions means that for a consensus of eight votes for the *weakest link*, there were eight ways for each

position and nine positions (A to I) giving 72 of the 132,217,728 million permutations that give a consensus of eight, with probability  $P(\text{consensus } 8) = \frac{72}{8^9}$ .

Following this line of reasoning for a consensus of seven, such that seven of the contestants voted for the Weakest Link in position A: -

Suppose B disagreed...

B can vote 7 ways (not B, not A),

A can vote 8 ways (not A),

There were 8 positions for disagree,

It gives,  $7*8*8$ .

Similarly, if 6 contestants voted for the Weakest Link in position A, and

B, C disagree...

B can vote 7 ways (not B, not A)

C can vote 7 ways (not C, not A)

A can vote 8 ways (not A)

there were 8 positions for disagree 1

there were 7 positions for disagree 2

As there were repeats (e.g., C, B disagree is the same as B, C disagree), divide by 2.

It gives,  $\frac{(7*7)*8*(8*7)}{2}$

From this, the formula below was derived for the number of permutations,  $P$ , for any consensus of  $n$ , where  $0 < n < 8$ .

$$P(n) = 7^{(8-n)} 8 \left( \frac{\frac{8!}{n!}}{(8-n)!} \right)$$

This formula was used to generate the permutations in the second column of Table 6 below. Notice that the permutations sum to the total number of possible permutations  $8^9$ . In the cases when the consensus was 8, 7, 6 or 5, these permutations also yielded contestant A as the *weakest link* by majority vote (the third column and tenth column in Table 6).

However, when consensus was between 1 and 4, the identity of the *weakest link* had to be enumerated for the cases where either another contestant received a majority vote or when there were ties. For example, when there was a consensus of four votes for contestant A as the *weakest link*, there was also the possibility that another contestant (e.g., B) received a majority vote of five. Also, there was the possibility that there would be a tie with another contestant also receiving four votes. The number of ties involving a consensus vote of four was 10,640 (column four in Table 6), all of which were 2-way ties resulting in  $\frac{10,640}{2} = 5,320$  permutations giving contestant A as the *weakest link* (Column five in Table 6 below).

The total number of permutations for generating contestant A as the *weakest link* with a consensus of four votes was therefore the sum of the permutations giving contestant

A by majority plus the ties  $1,333,640 + 5,320 = 1,338,960$  (column nine in Table 6 below).

This gives a probability of  $\frac{1,338,960}{14,913,081} = .0898$  (Column ten in Table 6 below) for a contestant to be voted as the Weakest Link with a consensus of four. Table 6 shows the relative frequencies and probabilities that would be expected purely due to chance for each of the different levels of consensus. This formed the basis for the null hypothesis for the distribution of consensus votes.

Table 6

*The Calculated Frequencies for Null Consensus for Different Level of Consensus*

Consensus	Permutation	Majority	Ties	2-way	3-way	4-way	9-way	Total - Ties	Probability
0	46,118,408	-	-	-	-	-	-		
1	52,706,752	-	133,496	-	-	-	14,833	14,833	0.000994623
2	26,353,376	526,568	16,628,304	2,726,472	2,988,160	552,720	-	6,793,920	0.455567837
3	7,529,536	5,840,912	1,523,760	754,320	5,040	-	-	6,600,272	0.442582726
4	1,344,560	1,333,640	10,640	5,320	-	-	-	1,338,960	0.089784265
5	153,664	153,664	-	-	-	-	-	153,664	0.010303974
6	10,976	10,976	-	-	-	-	-	10,976	0.000735998
7	448	448	-	-	-	-	-	448	3.00407E-05
8	8	8	-	-	-	-	-	8	5.36442E-07
Total	134,217,728	7,866,216	18,296,200	3,486,112	2,993,200	552,720	14,833	14,913,081	1

**Note.** See text above for explanation.



Table 7

*The Calculated Frequencies for Voting by Consensus (72 – episode Study)*

Consensus	72-episode	Null
1	0	0.071612835
2	1	32.80088425
3	12	31.86595631
4	16	6.464467049
5	17	0.741886139
6	11	0.052991867
7	12	0.002162933
8	3	3.86238E-05
Total	72	72

**Note.** Shows the calculated frequencies by consensus for 72-episode and based on the null hypothesis probability (refer to Table 6).

The comparison of consensus frequencies between the observed 72-episode study and the null hypothesis distribution of votes is shown in Table 7 above and plotted in Figure 14 below. Figure 14 shows the vote frequencies cast by contestants as a function of consensus based on the probability, ( $\chi^2(2, N = 72) = 411.924, p < .001$ ; large effect size, Cramer's  $V = 1.691$ ). Notice that the observed pattern differs from that expected due to chance. This is not a surprise as the contestants taking part in the actual game-show have the added information available of observing the game play of their fellow contestants and are thus less reliant on casting random votes.

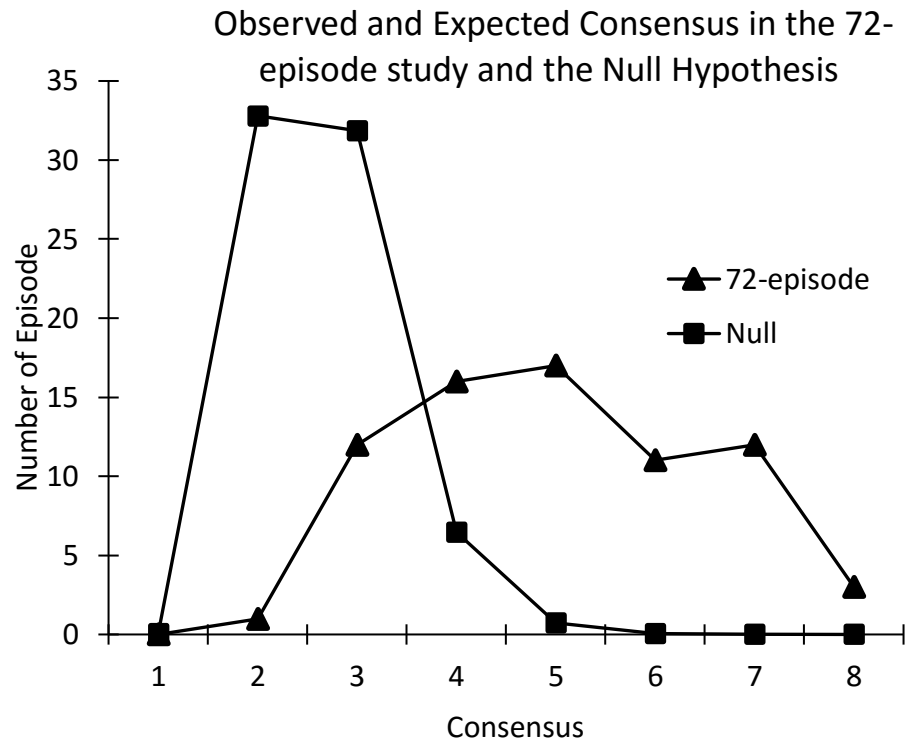


Figure 14. shows the consensus based on probability (squares) and votes cast by contestants (triangles) by consensus level. See Table 7 for the consensus frequency and probability values used to generate the null hypothesis.

*Partitioning Consensus.* The overall pattern of voting across the 72-episode study was described previously by Figure 10.1 and 10.2 above. It revealed three distinct phases of voting behaviour at three different levels of proximity between the voter and votee; proximal, medial, distal. The proximal phase incorporated the NE and only included the most proximal voter-votee relationship between the voter and their immediate neighbour(s) (N). The NE was characterised by a positive ratio  $\frac{144}{115} = 1.252$ . The medial (N+1, N+2, N+3) and distal (N+4, N+5, N+6, N+7) phases had respective vote ratios of 0.904 and 1.137. The three phases having been characterised as ratios provide an indication of the strength of voting bias across the three phases as plotted in figure 15 below. A ratio greater

than one shows the NE as a bias in the proximal phase, as the votes observed were less than the expected vote. When the ratio was less than one in the medial phase, the observed votes were more than expected. A ratio of one occurred when the observed votes were as expected.

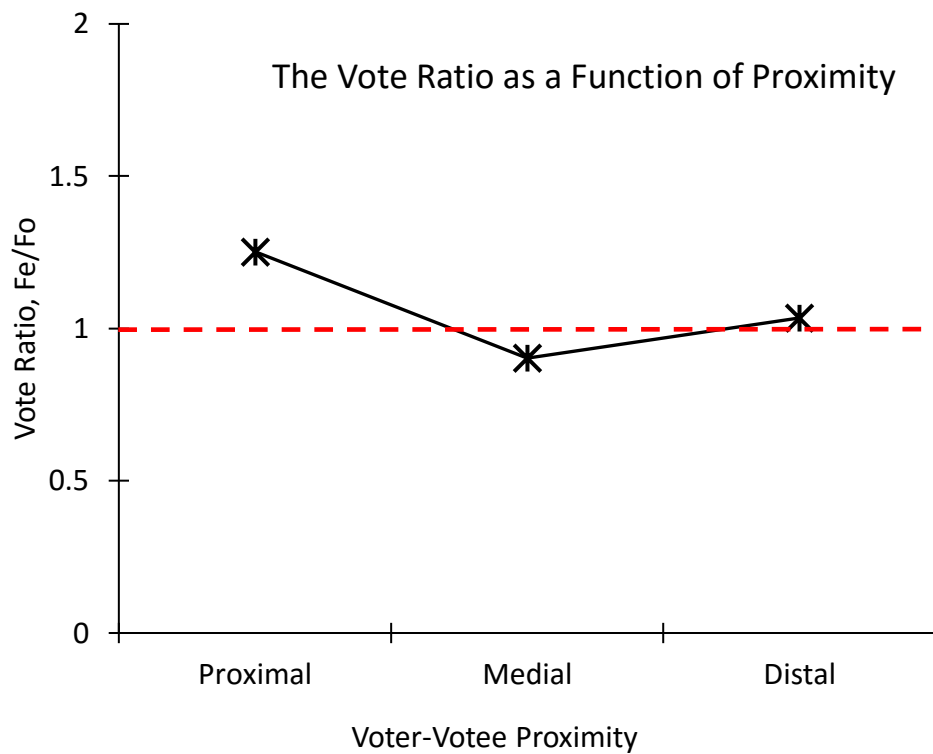


Figure 15. shows the vote frequency ratio as a function of spatial proximity for the 72 – episode study. The proximity is divided into three phases (*proximal*, *medial*, and *distal*). The red line indicates the ratio is equal to one as the baseline.

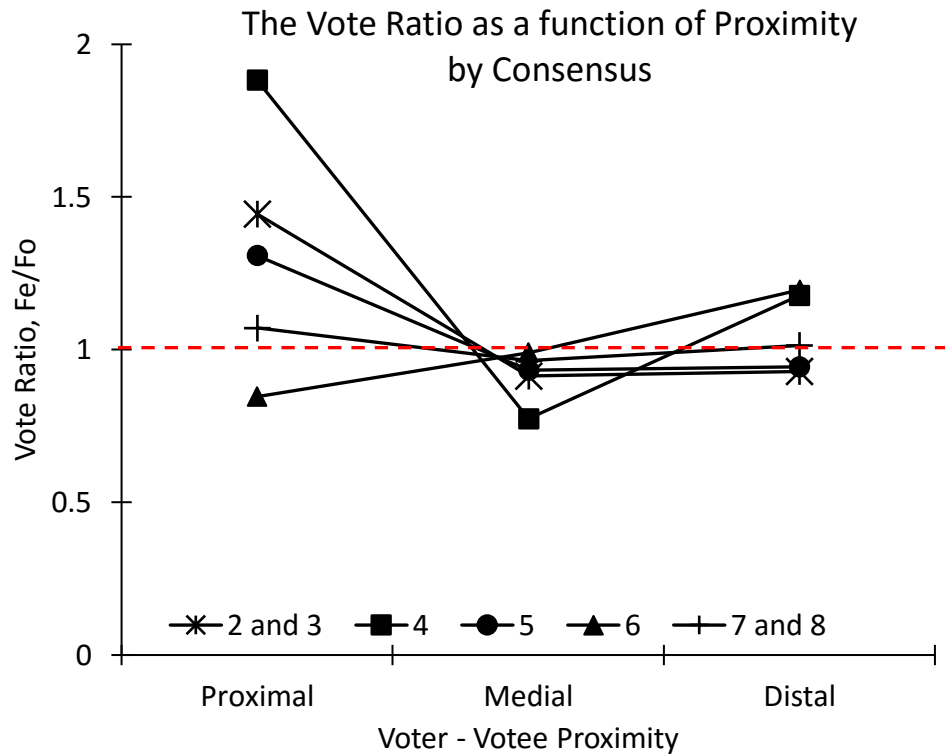
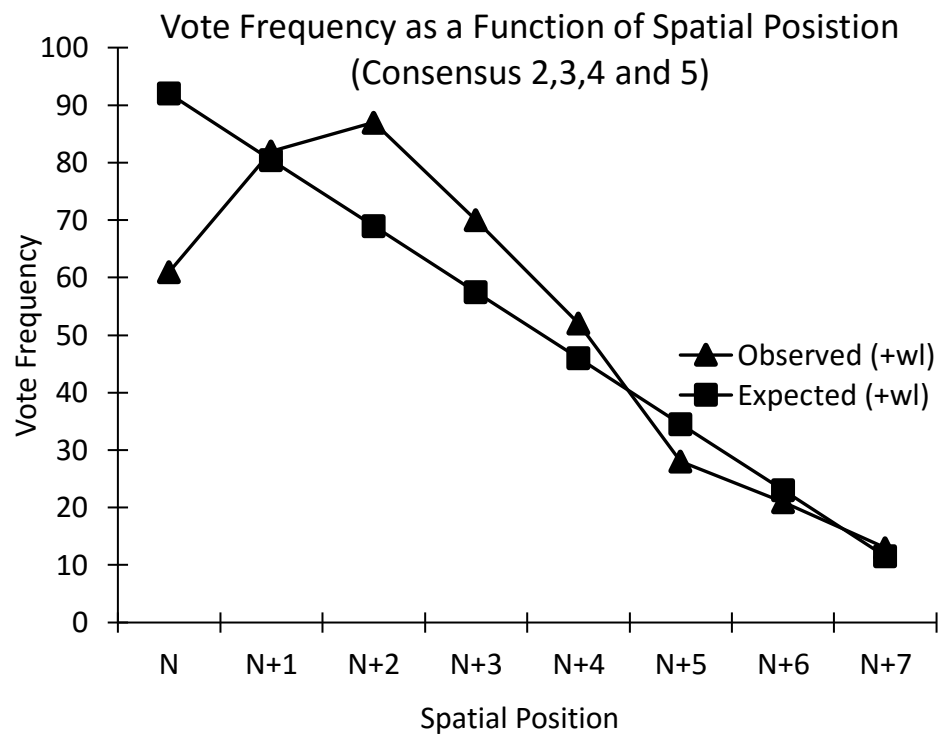
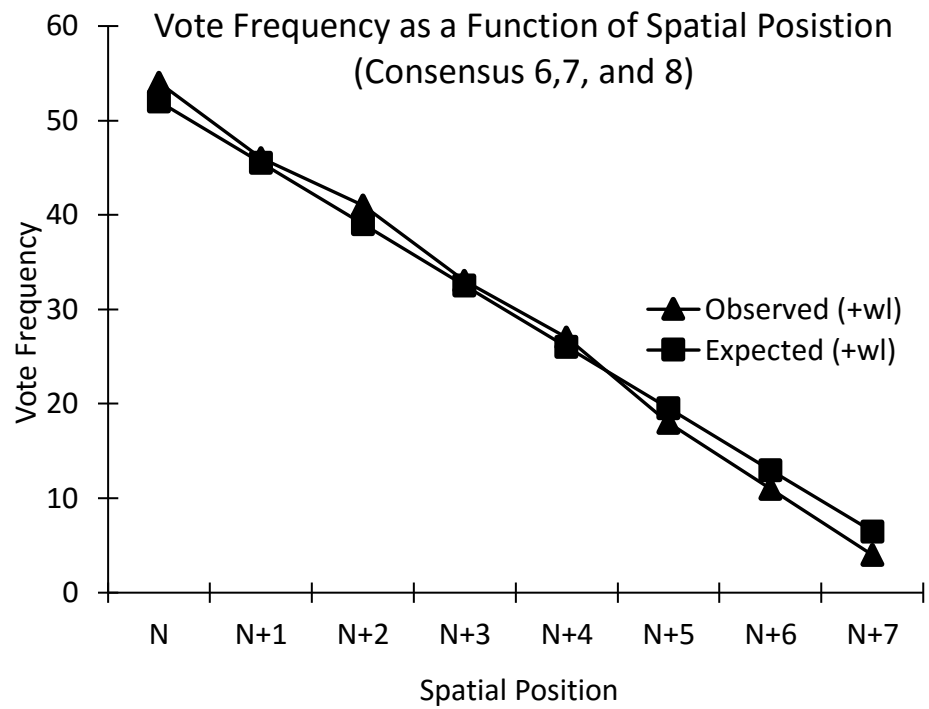


Figure 16. shows the same data as in Figure 15 above using the same axes. In this figure the consensus was partitioned from low to high (2/3, 4, 5, 6, 7/8). The ratio was measured by dividing the expected vote frequency for each consensus with the observed vote frequency. Notice that the NE in the proximal phase tended to be greater for lower consensus.

Figure 16, above, shows that at the *proximal* phase of proximity, at lower consensus levels (2/3, 4, 5) the ratio was greater than one indicating a NE bias. The NE (at *proximal* proximity) was strongest when the voting consensus was 4, 1.882, nearly half less votes than expected. This was followed by the consensus at 2/3, 1.444, and consensus 5, 1.308. In contrast when consensus was highest (7/8) the ratio was just above one, 1.071, and even dropped below one at a high consensus of 6. It shows that those episodes with lower consensus (2/3, 4, 5) show a strong NE and episodes with high consensus show no

NE (6, 7/8). The remainder of this section compares voting between these high and low consensus episodes.

Figures 17.1 and 17.2 below, show that the voting pattern changes by partitioning consensus. Figure 17.1 (upper panel) shows *high* consensus (n=26 episodes; combining episodes where 6, 7 or 8 contestants voted for WL) because a large number contestants agreed on the WL, then there was no observable NE because the observed voting pattern was largely as expected ( $\chi^2(7, n = 234) = 1.616, p = .978$ ). However, once the consensus as to the identity of the WL dropped, a marked NE emerged when the observed vote frequency dropped well below the expected frequency at N ( $\chi^2(7, n = 414) = 20.263, p = .005$ ; Cramer's  $V = .084$ , small effect size) (see Figure 17.2, lower panel). When consensus was *high* the voting pattern was as expected but when *low* there was a notable bias in the voting pattern showing the characteristic proximal, medial and distal phases, with a strong NE at the proximal phase.



Figures 17.1 and 17.2. shows the voting pattern at *high* (at 6, 7 and 8) and *low* (2,3,4 and 5) consensus episodes.

**Discussion.** The main purpose for undertaking this reanalysis of Goddard et al. (2011) data was two-fold. Firstly, to become acquainted with how to code and analyse data from the observational field study and secondly to carry out a replication of their original findings. The main finding from Goddard et al was largely confirmed, that contestants show a strong avoidance for voting for their direct-neighbours but that the voting was not a simple function of the spatial distance between the *voter* and the *votee*. As the consensus regarding the weakest link reduced then the NE in the contestants' voting increased. Additionally, however, this replication undertook an analysis of voting by gender and using the GBS of Anotonvics et al. (2005) revealed another strong bias regarding gender where both male and female contestants show significant discrimination by voting against women.

**TYPE I v TYPE II info/uncertainty.** One possibility for how a contestant made a decision was that they made their voting decision based on the availability of two very different sources of information. Firstly, TYPE I *game-specific* information (i.e., exogenous factor) was the primary source of information. This kind of information was public, explicit and encompassed the observable performance of the contestants during the round of questions. The kinds of factors that could contribute towards this would be the measurable game performance of the other contestants encompassing aspects of game-play like the accuracy of answers provided and the accuracy of banking decisions made.

However, the cognitive task of keeping track of the eight other contestants' performance in answering something in the region of 30 questions in a three-minute interval is a challenging one under the best of times. Managing this task whilst also taking

part in answering the questions and being subject to the peculiar demands of being in the high pressure environment of the TV studio makes making their voting decision an even more daunting prospect. Rather than keeping some accurate tally of all of the other contestants it is far more likely that they register the more distinctive and observable negative aspects of their fellow contestants' performance. This would be in the form of obvious errors or even other perceived gaffes like errors in answering questions especially if allied with a failure to bank a large sum. When once the integrity of this source of information was high and unambiguous, contestants faced a relatively easy and reliable decision identifying the weakest performer. This is particularly the case in the 'high' consensus rounds where there is near unanimity over the identity of the weakest link.

On the other hand, in rounds when *game-specific* information becomes equivocal then *voters* were in a quandary on who to pick as weakest link. For example, in episodes when all candidates answered all of the questions correctly, or, in episodes where many of the contestants performed badly, it is unclear who the weakest link really is. In these cases, where TYPE 1 information is worthless so, cannot be used, they had to rely instead on secondary *subject-dependent* sources of information. This TYPE II *subject-dependent* (i.e., endogenous factor) source was private, implicit and encompassed each voter's individual subjective dependent attributions, discriminations and other biases. In these rounds, group consensus over the identity of the weakest link was lower because decision-making shifted away from the primary high fidelity source of *game-specific* information to the secondary *subject-dependent* source, more prone to biases like NE and the other forms of implicit discrimination that were the subject of Levitt's (2004) and Antonovics et al. (2005) research. The increase in the NE as consensus dropped below six highlights the



increase in uncertainty as to the true identity of the weakest link. It would also be expected that these kinds of discrimination would come to prominence during the first-round episodes where there was low consensus about the WL's identity.

***Gender discrimination.*** The most dramatic addition to the analysis conducted by Goddard et al. (2011) was the finding of a strong gender effect in the voting of all contestants against women contestants. What remains uncertain is the cause of this apparent discrimination. Discrimination can take different forms.

The preference-based discrimination in the first-round voting by females against males was attributed by Antonovics et al. (2005) as a preference for women rather than necessarily as a dislike for men. However, the outcome amounts to the same which was a negative outcome for males that was otherwise undeserving. This discrimination disappeared immediately after the first-round. Statistical discrimination could arise if the group performance was statistically worse for men than women. This was not the case in Antonovics et al. data. Males performed effectively the same as females regarding the accuracy of their answers to questions. Antonovics et al., raised a second kind of statistical discrimination, where the signals on ability are more informative for one group than another. Typically, though, the signaling is stronger for the in-group so would be applied more to women-on-women discrimination. A third possibility is incorrect prior beliefs that men perform worse. However, after a number of rounds this would be corrected. Strategic alliances in the form of explicit/implicit group collusion was another possibility raised although Antonovics et al. found little evidence for this.

The discrimination they found is still open to interpretation. They conclude that “There are a number of possible explanations for this type of behaviour” (Antonovics et al., 2005, p. 942). The discrimination shown is considered to be preference-based largely on the basis that this is the remaining form of discrimination that cannot be discounted. Women might have some kind of dislike for some aspect of how men play the game. They might have more compassion for other women, or even they might wish to avoid competing against men in future rounds. Gneezy, Niederle and Rustichini (2003) found that in experimental settings women perform worse when competing against men. Women also seem to favour women when voting for political candidates Dolan (1998). While, Levitt (2004) talks of difficulty addressing discrimination.

The discrimination found here is more difficult to address because unlike Levitt (2004) and Antonovics et al. (2005), Goddard et al. (2011) data coding did not include the performance of the contestants or any other demographics aside from contestants’ gender, position on the game-show (A-I), their first-round vote choice. As the original episodes are no longer available, it is not possible therefore to ascertain whether the women that the contestants voted for was based on their game playing performance or their membership of the group identified as “female”. Women in the U.S show a preference for other women whereas the women in the UK show discrimination against women. It remains to be seen whether these reflect cultural differences or some aspect of their game playing performance.

The only way to resolve the issues relating to this replication of Goddard et al. (2011) is to carry out a further replication by repeating the field observation of the WL

using a different set of episodes including a consideration of the game performance of the contestants. The aim of the next study is to carry out this replication with an emphasis on testing for voting bias in proximity/gender/game performance and consensus to investigate the endogenous and exogenous factors in the voting behaviour.

### R1.3 Observational Field Study 2: Replication of The Neighbour Effect

**Abstract.** The main question for this research was to verify whether the first-round voting behaviour previously observed in the preliminary study was affected by the *voter-votee* physical proximity. The analyses of results are presented in three parts; the *neighbour effect* (NE), the gender effect and consensus. Voting behaviour, position gender and game performance were recorded for 1,359 contestants (Female = 606, Male = 753) in 151 episodes of the UK version of the WL TV game-show. The main findings of Goddard et al. (2011), and the reanalysis above, were supported because there is a strong significant voting effects were recorded for *voter-votee* proximity and consensus. The other main finding was that the gender effect found in the reanalysis of Goddard et al. was not apparent in this extended data set, instead, traces of discrimination were much weaker and had reversed. Whereas, the data reanalysed from Goddard et al showed a strong discrimination against women, the data here showed a weaker discrimination against men.

**Introduction.** From the preliminary study that reanalysed Goddard et al. (2011) data there was a significant association between proximity and voting behaviour. It showed the NE because the proximity bias was strongest towards contestants who were physically closest to the *voter*. On top of that, there was gender discrimination where both genders voted against females as first-round weakest link. Therefore, this study was carried out using more episodes to test, if both effects would still be evident using a larger set of episodes.

**Method.** The method used was for the most part the same as that described above and relayed in the General Methods chapter **M1.2**.

Data was obtained by observing 151 episodes of the UK version of WL TV game-show broadcast on UK free to air TV in the United Kingdom between 1<sup>st</sup> May 2009 and 17<sup>th</sup> January 2011. The observation excluded the celebrity episodes. In each episode, there was nine contestants, hence a total of 1,359 contestants' (Female = 606, Male = 753) voting activities were observed. Other demographic factors (i.e., age, marital status, ethnicity, education, income and occupation) were not recorded due to limited information and it was not applicable for the study.

A simple database was used to code data collection in Microsoft Excel. Information recorded included the file name, episode date, spatial position of each contestant (A-I), their gender, their vote and additional remarks for any observed behaviour. The collected data was cross-checked to ensure the collected data was correct with no error. Computation of the accumulation of the WL's votes were computed by using a simple function in the Excel formula builder and the data was validated by another independent researcher for the data authenticity. All of the observations were made by the principal researcher and author of this thesis.

**Results.** The analysis used in this section always followed the same pattern as that described above in the replication of Goddard et al. (2011). Again, the *observed* frequency of votes actually made by contestants was counted and compared with the frequency of votes *expected* purely due to probability (Goddard et al.). The method used to explain how

the *expected* frequencies were derived was explained at the beginning of the results section in the previous study but is briefly explained again here using 151 episodes. The structure of this results section also follows the same path as that from the previous study. The first replicated the analysis of Goddard et al. regarding voting behaviour and proximity, the second part tests for the gender effect exposed by the previous study and the final part examines proximity and consensus.

***Calculation of expected vote frequencies in 151 episodes.*** Table 8 below, shows how the expected frequencies for a single episode were simply multiplied by the number of episodes, 151, to generate the expected vote frequencies for each *voter-votee* spatial relationship in the first-round.

Figure 18 below, shows the expected frequencies from Table 8 plotted as a function of spatial position. Error bars for each expected frequency are estimated from the variance ( $\pm 2$ ) of the underlying theoretical probability distribution,  $\frac{pq}{N}$ , with,  $p$ , the expected percent of votes for that spatial position, and  $q = (100 - p)$ .  $N$  was the total number of votes made.

The votes from each contestant ( $N = 1,359$ ) were recorded based on the relative spatial positions of *voter-votee*. Observed and expected vote frequencies were calculated to both include and exclude votes from the first-round WL. Recall that a vote denoted the negative choice that been made by the *voter* to a *votee*.

Table 8

*The Calculated Expected Frequencies for 151 Episodes*

Pos.	1 ep.	151 ep.
$N$	2.000	302.000
$N+1$	1.750	264.250
$N+2$	1.500	226.500
$N+3$	1.250	188.750
$N+4$	1.000	151.000
$N+5$	0.750	113.250
$N+6$	0.500	75.500
$N+7$	0.250	37.750

**Note.**  $N$  is one space away neighbour, hence +1, +2, +3, ...+7 denoted as added spaces in addition to  $N$ . For every calculated frequency, each spatial position was multiplied with number of episode (ep.) observed.

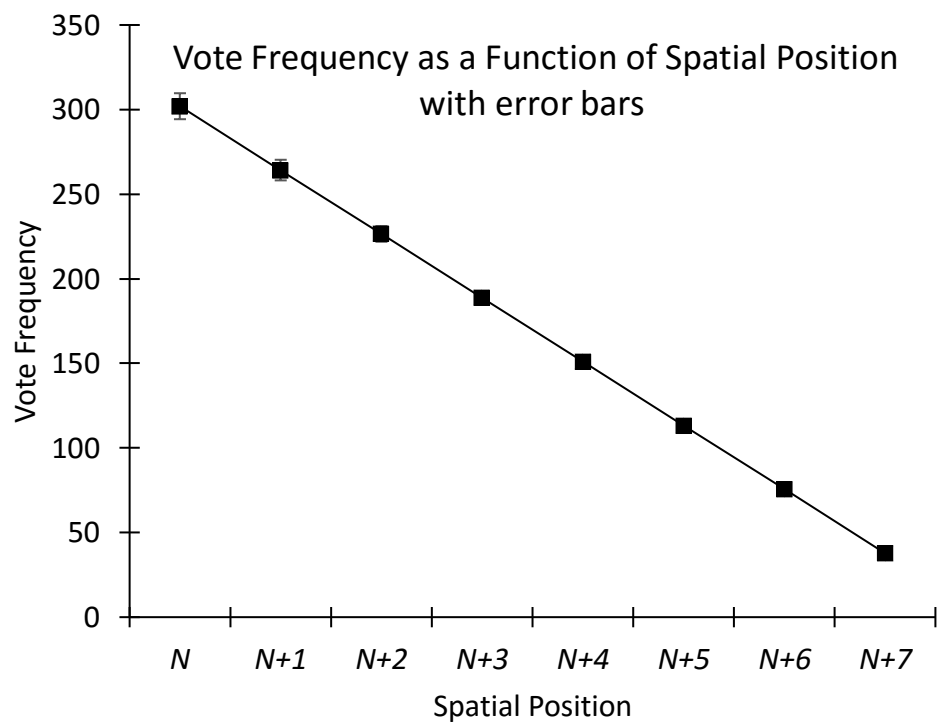
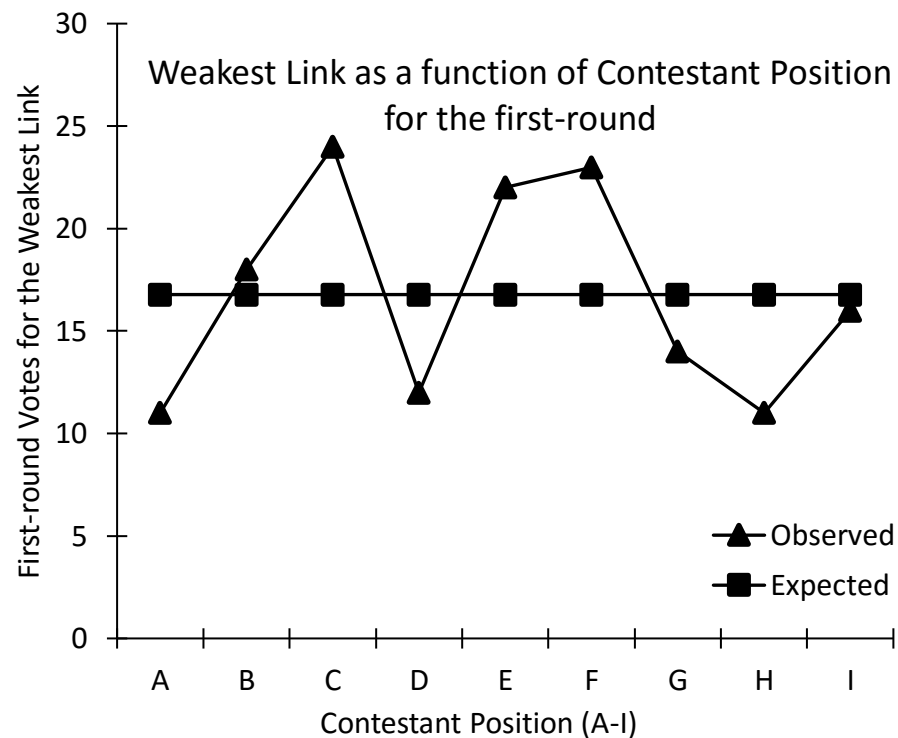


Figure 18. shows the expected voting pattern with error bars (see text above).

***WL as a function of contestant position.*** The position of the weakest link throughout the 151 episodes was recorded and the frequencies that first-round weakest links appeared in positions A-I is shown in Figure 19. Over the course of 151 episodes there should be no significant deviation from the expected 16.778 occurrences of the weakest link in each position ( $\frac{151}{9} = 16.778$ , triangular symbols). There was no systematic bias in the spatial or temporal order of candidates as the observed frequencies ( $f_o$ ) of the weakest link found in each spatial position (A - I) compared to the expected frequency ( $f_e$ ) of 16.778 ( $n(\text{episodes})/n(\text{positions}); \frac{151}{9} = 16.778$ ) was non-significant ( $\chi^2(8, N = 151) = 12.967, p = .113$ , small effect size, Cramer's  $V = .104$ ). Contestants were no more or less likely to vote for candidates that preceded or followed them temporally in the testing sequence.



*Figure 19.* Shows the distribution of the WL across the nine contestant's position A - I for the first-round (151-episode study).



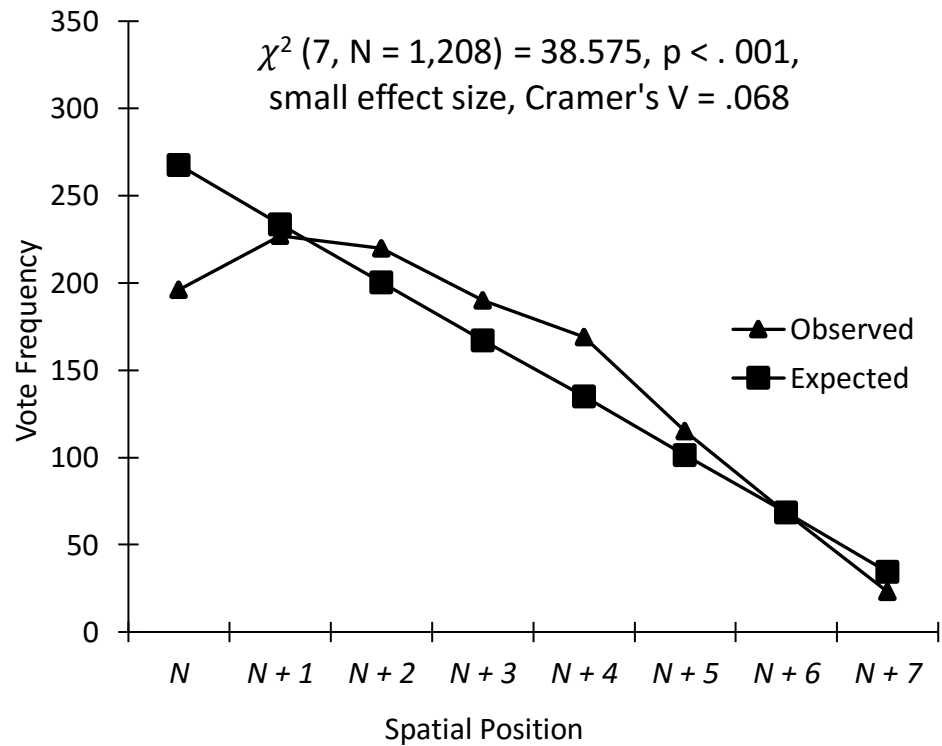
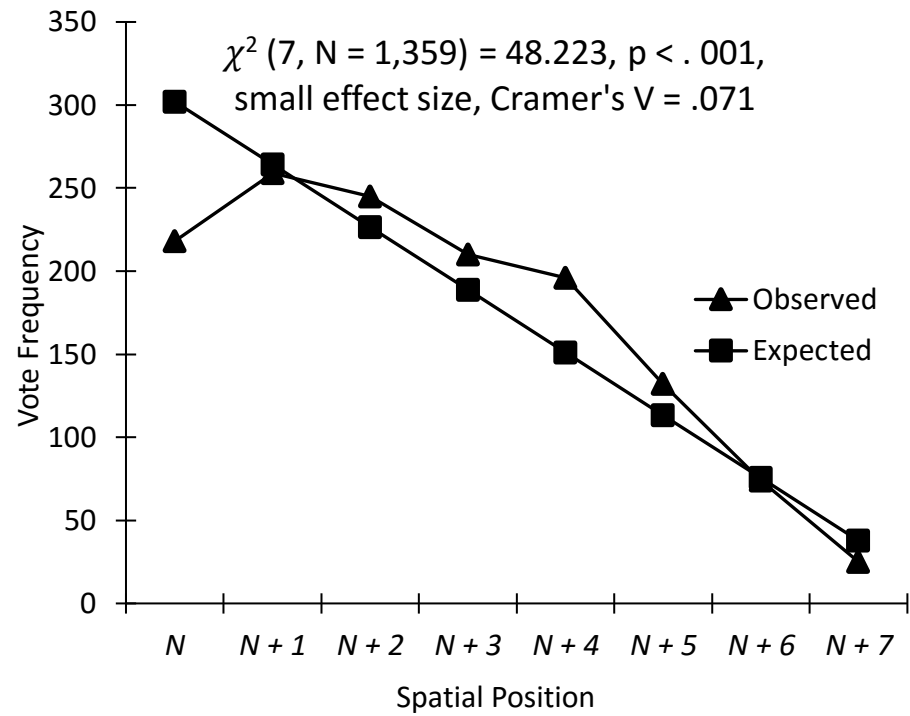
**Finding 1: Voter-votee proximity (The “neighbour-avoidance” effect).** The calculated  $\chi^2$  for each spatial position over 151 episodes is shown in Table 9. Figures 20.1 and Figures 20.2 show the comparison between the observed and expected voting frequency as a function of voter-votee proximity. The observed frequencies show the now familiar three-phase *proximal-medial-distal* pattern of voter-votee first-round voting. Again, there was no simple linear relationship between votes and proximity. The result shows that, there is a significant association between spatial position and the vote frequencies of *WL*,  $\chi^2(7, N = 1,359) = 48.223, < .001$  small effect size, Cramer’s  $V = .071$  (The data shown here includes the votes cast by the eventual weakest link, see Figure 20.1, the upper panel). It shows that there is a significant descent at the  $N$  point, which is the direct spatial contestant. This demonstrates the proximity effect even more clearly than in Goddard et al. (2011) 72-episode study. Therefore, the null hypothesis was rejected and  $H_1$  was accepted. Figure 20.2 (lower panel) demonstrate a significant NE,  $\chi^2(7, N = 1,208) = 38.575, p < .001$ , small effect size, Cramer’s  $V = .068$  (The data shown here excludes the votes cast by the eventual weakest link).

Table 9

***The Chi-Square Value for Each Spatial Position.***

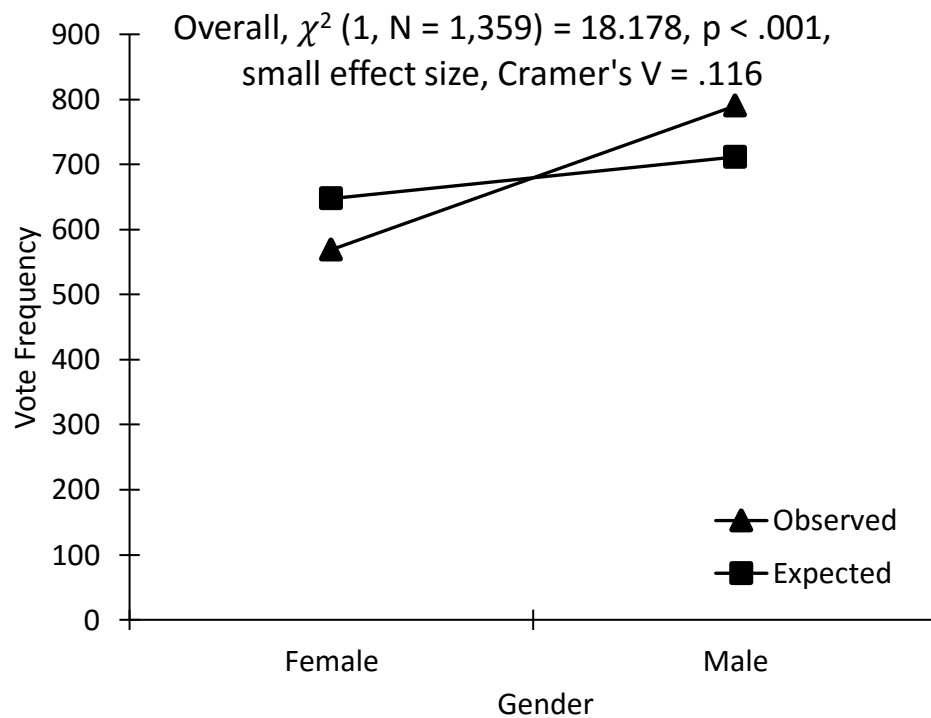
Spatial position	Observed	Expected	$\chi^2$
$N$	218.00	302.00	23.36
$N+1$	259.00	264.25	0.10
$N+2$	245.00	226.50	1.51
$N+3$	210.00	188.75	2.39
$N+4$	196.00	151.00	13.41
$N+5$	132.00	113.25	3.10
$N+6$	74.00	75.50	0.03
$N+7$	25.00	37.75	4.31
Total	1359.00	1359.00	48.22

**Note.** The observed and expected frequencies were derived from 151 episodes of the WL,  $N = 1,359$ ,  $df = 7$  (The data shown here includes the votes cast by the eventual weakest link).



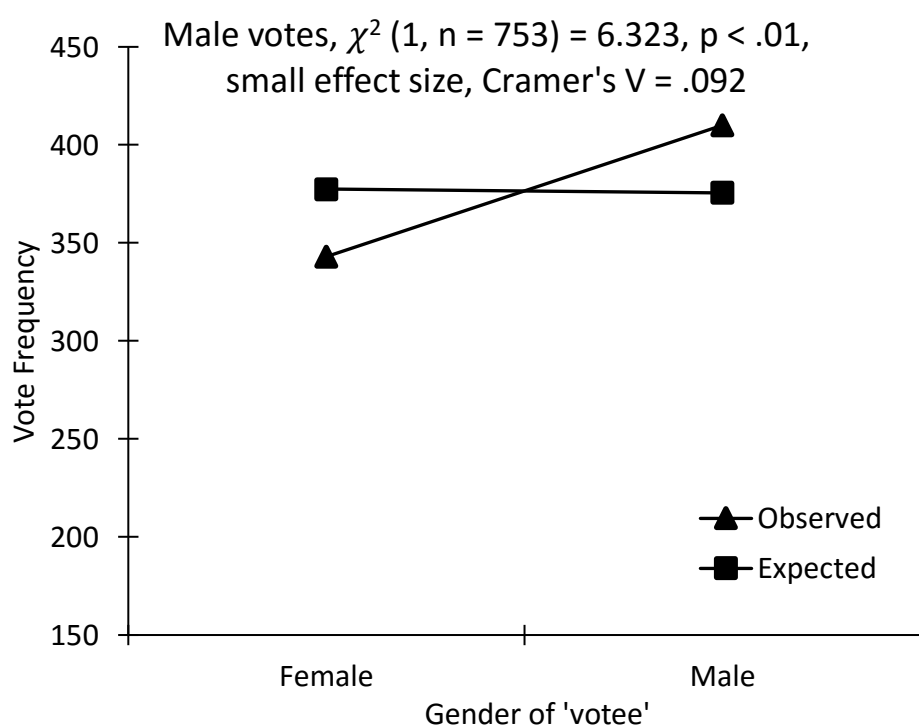
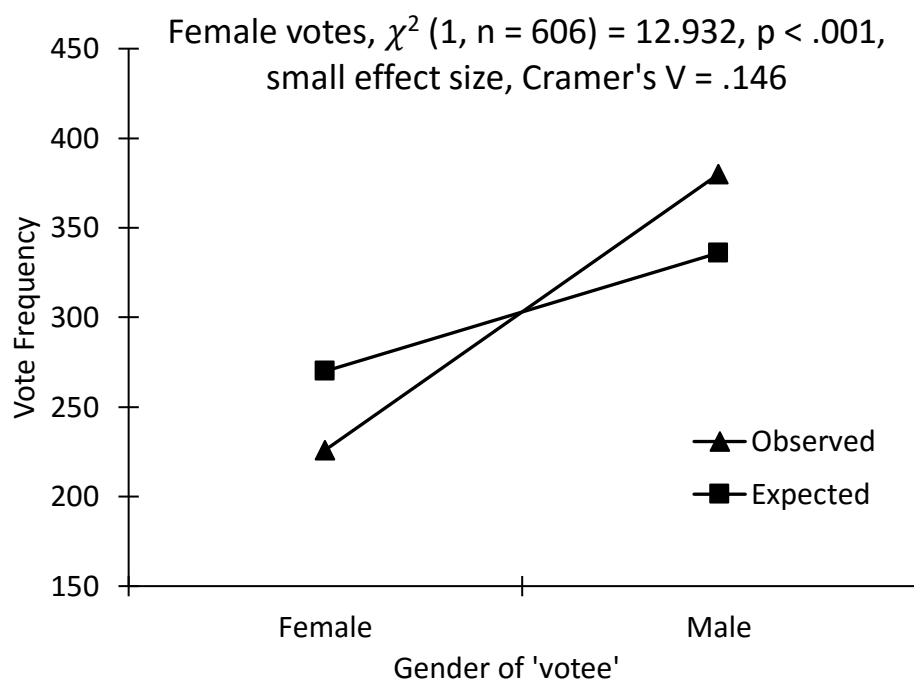
*Figures 20.1 and 20.2.* Compares the expected voting frequency, derived as the unbiased pattern of voting, with the observed pattern of voting. Voting frequency is plotted as a function of the spatial relationship between voter and candidate where  $N$  refers to the voter's direct-neighbour and  $N + 1$  the next-door-but-one neighbour and so on.

**Finding 2: The effect of gender.** A simple analysis in the 72 episodes' study revealed a strong gender bias regarding the identity of the weakest link in the first-round voting (i.e., 46 Females versus 26 Males). At least for the first-round of voting there was a marked discrimination in voting. In the extended observations over 151 episodes this first-round gender effect disappears. The expected frequencies for weakest link (Female = 63.336; Male = 83.667) compared closely to the observed weakest link (Female = 61; Male = 90). If anything, the bias reversed from the 72 episodes' study but there was no significant departure from the expected frequencies in the 151 study ( $\chi^2(1, N = 151) = .978, p = ns$ ).



*Figure 21.* Shows the vote frequencies with respect to female and male contestants for 151-episode study

Breaking these results down further, Figure 21 shows the observed and expected vote frequencies for female and male voters voting for females and male respectively. In the reanalysis of Goddard et al. (2011) data females in the first-round of voting were subject to significant negative bias and discrimination by both male and female voters in the 72 episodes' study. Figure 21 shows that this pattern of discrimination has changed completely in the replication using 151 episodes. Figure 21 shows the overall expected votes for males and females and the actual observed frequencies,  $\chi^2(1, N = 1,359) = 18.178, p < .001$ , small effect size, Cramer's  $V = .116$ . The votes targeted towards women were significantly fewer than expected with the corollary that male votes were greater than expected. Figures 22.1 and 22.2 show the separate votes made by males and females. Figure 22.1 (upper panel) is a voting performance for female contestants,  $\chi^2(1, n = 606) = 12.932, p < .001$ , small effect size, Cramer's  $V = .146$  and Figure 22.2 (lower panel) shows a voting performance for male contestants,  $\chi^2(1, n = 753) = 6.323, p < .01$ , small effect size, Cramer's  $V = .092$ .



Figures 22.1 and 22.2. Voting performance for both male female contestants. Figure 22.1 (upper panel) is a voting performance for female contestants and Figure 22.2 (lower panel) is a voting performance for male contestants.

Women now show the kind of discrimination also found by Antonovics et al. (2005). Females show a preference for other females by being more likely to vote for males. Males, show a similar bias because they also avoided voting for females at the expense of males. Updated GBSs for the 151-episode study gives females = 0.991 (n = 606) and males = 1.092 (n = 753). Goddard et al. data yielded a GBS of 1.21 (n = 300) for females and 0.764 (n = 348) for males.

*Distribution of men and women.* The arrangements by gender for each episode were observed and in total five arrangements were identified. Table 10 shows the different arrangement by gender for the 151 episodes. The commonest format for WL TV game-show was the same as used in the original 72-episode study (Male-Female-Male-Female-Male-Female-Male-Female-Male). Though in the 151-episode study the arrangement accounted for 93.378% of episodes relative to the 40.278% of episodes in the 72-episode study. Only five arrangements were encountered in the 151-episode study compared to 25 different arrangements in the 72-episode study. Of the 72 episodes, 60 had five males and four females with the remainder having five females and four males. Of the 151 episodes, 149 had five males, only two had five females with four females.

Table 10

Contestant Arrangement by Gender (151- episode study)

<u>Arrangement</u>	<u>Number of episodes</u>
mfmfmfmfm	141
mfmfmfmfmf	6
fmfmfmfmf	2
fmfmfmfmfm	1
mmfmfmfmf	1

Female and male contestants voted males more than females as WL which rebuts the finding in the previous 72-episode study where the observation showed a significant gender bias against other females. These results are more in line with previous researchers using WL (Antonovics et al., 2005; Levitt, 2004).

***Finding 3: The effect of consensus.*** The assumption was made that when all eight contestants (high consensus) voted the same contestants as weakest link, it indicated a high level of certainty to identify the weakest contestant. Similarly, a low consensus was indicative of uncertainty identifying the WL. The prediction is that the NE increases with uncertainty. The distribution of the voting consensus among the contestants over 151 episodes is shown in Table 11. It shows that the commonest consensus was when all eight contestants were unanimous by voting for the same WL.

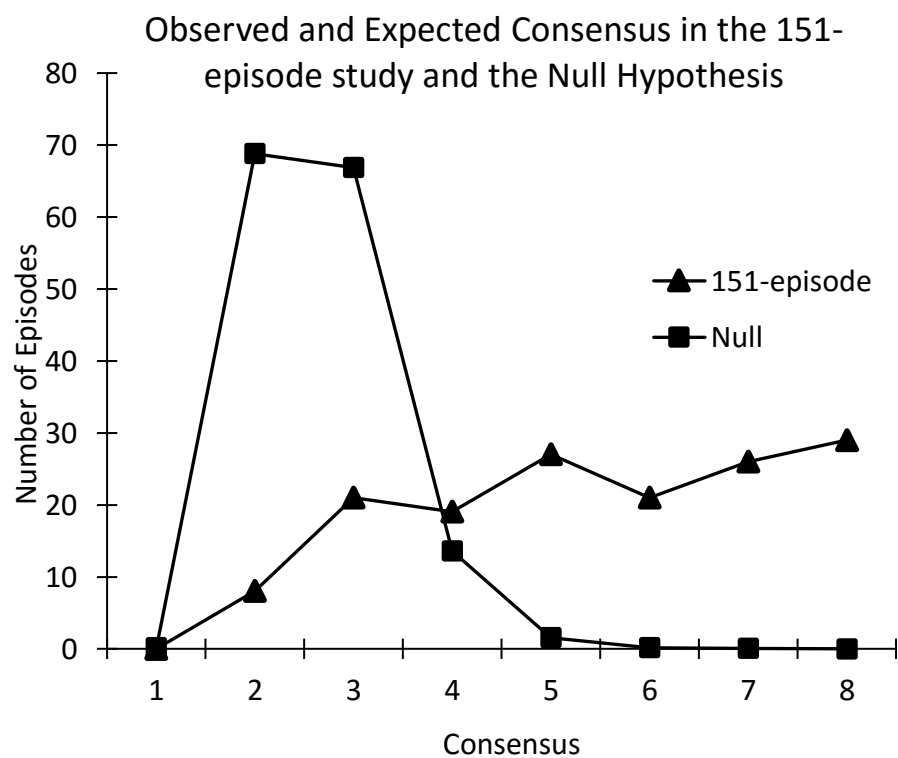
Table 11

*The Contestants Consensus Votes in 151 Episodes*

Consensus	151-episode	Null
1	0	0.150188029
2	8	68.79074335
3	21	66.8299917
4	19	13.55742395
5	27	1.555900097
6	21	0.111135721
7	26	0.004536152
8	29	8.10027E-05
Total	151	151

**Note.** Shows the calculated frequencies by consensus for 151-episode and based on the null hypothesis probability (refer to Table 6).

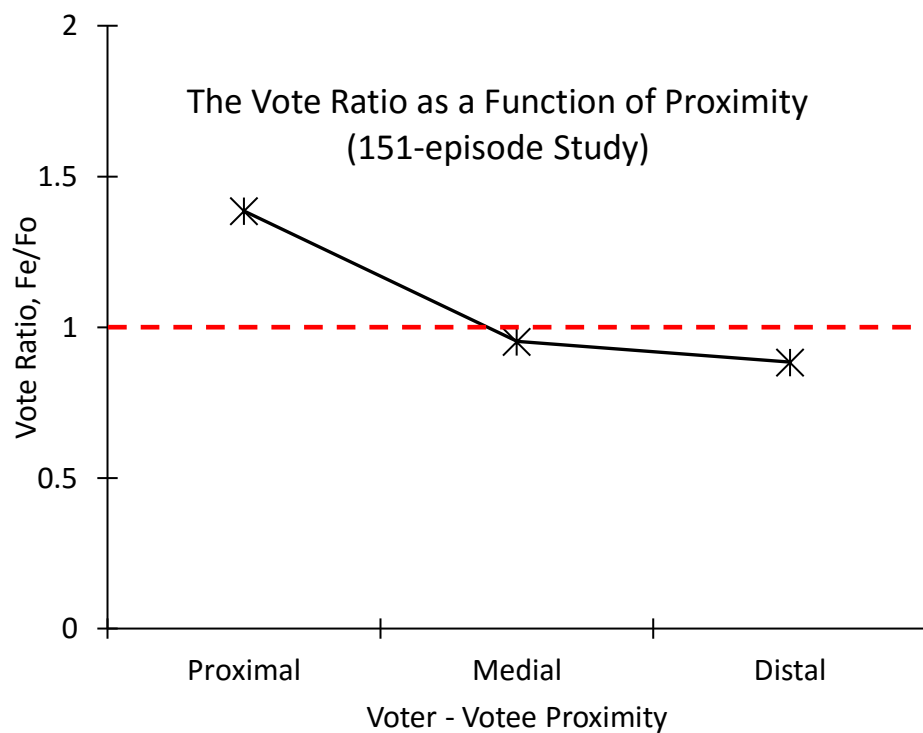
The comparison of consensus frequencies between the observed 151-episode study and the null hypothesis distribution of votes is shown in Table 11 above and plotted in Figure 23 below. Notice that in the actual episodes observed voting tended towards the highest consensus over the WL with the most frequent being the maximum consensus of eight ( $\chi^2(2, N = 151) = 833.868, p < .001$ ; large effect size, Cramer's  $V = 1.662$ ). This is not a surprise as contestants were likely in most cases to base their voting decision on the obvious poor performance of the worst contestant. In the random model derived from probability theory the modal consensus was a low two.



*Figures 23.* shows the consensus based on probability (squares) and votes cast by contestants (triangles) by consensus level for 151 episodes.



*Partitioning Consensus.* The overall pattern of voting across the 151-episode study was described previously by Figure 20.1 and 20.2 above. Similar to the 72-episode study, voting behaviour was grouped into three proximity phases (*proximity*, *medial* and *distal*). The grouping of the 151-episode study was done to give an indication of the strength of any voting bias as a function of proximity. Figure 24 shows that, at *proximal* proximity, the NE indicates a positive bias as the ratio given was more than one, at 1.385. Whilst, for the *medial* (.952) and *distal* (.884) proximity, the ratios were close to but less than one, which suggests that the contestants were showing either no bias, or a slight negative bias towards contestants medial or distal to them.



Figures 24. . shows the vote frequency ratio as a function of spatial proximity for the 151 – episode study. The proximity is divided into three phases (*proximal*, *medial*, and *distal*). The red line indicates the ratio is equal to one as the baseline.

**Discussion.** The main findings that came out from this replication were that they supported and strengthened the main conclusions of Goddard et al. (2011). Firstly, contestants showed significant bias in their first-round voting behaviour based on the proximal relationship between *voter* and *votee*, second, this proximity bias/effect was not a simple linear function of distance, third, three phases were apparent, proximal-medial-distal and fourth, the proximity effect was related to uncertainty, because as uncertainty increased with reduced weakest link consensus so did the proximity effect.

Using an extended data set of 151 episodes made the effects more clear cut than in Goddard et al. (2011) and in the reanalysis of Goddard et al above. Another main finding from this study was that the gender effect revealed in the reanalysis of Goddard et al., disappeared and even reversed in this study. The GBS for females of 0.991 was close to one indicating no discrimination. The associated  $\chi^2$  was significant however, showing a small effect size bias against voting men (similar to Antonovics et al. (2005) this can be attributed in part to the sensitivity  $\chi^2$  shows for high numbers of observations. Men showed a similar bias against other men by voting for them as WL.

A shortcoming of the reanalysis of Goddard et al. (2011) in the previous study was that it was based on already coded data without recourse to the original episodes making it impossible to determine the source, preference-based or statistical-based, for the discrimination against women discovered. The current replication had the advantages of being based on more episodes that the principal researcher had direct access to. This meant that the performance of contestants was recorded and it showed that there was no difference in the accuracy of men and women addressing first-round questions. Males

fielded first-round questions with an accuracy of 0.890 compared to females' accuracy of 0.885. Consequently, there is no performance-based statistical information to hand for contestants to favour either males or females. The discrimination that emerged, akin to that exposed by Antonovics, et al. (2005), was most likely preference-based. The precise explanation for this discrimination remains open to interpretations however. The difference between these observations and those reported by Levitt (2004) and Antonovics et al. was that both males and females here showed a preference-based discrimination for women, or, against men. Whichever, it amounts to a kind of discrimination.

It remains unclear, however, why there was such a marked discrimination against women in Goddard et al. (2011) data. One possibility is that their data incorporated statistical-based discrimination against women due to poor performance by women during the first-round, otherwise it is difficult to arrive at an explanation for why preference-based discrimination should shift so dramatically in the space of less than a decade between the recordings from the two studies.

## R1.4 Observational Field Study 3: Analysis of Second-round Voting Reciprocity, Retaliation and Consistency.

**Abstract.** The pattern of voting in the second-round of 72 episodes of the UK version of the WL TV game-show was analysed. The observations were based on the original data set used by Goddard et al. (2011). In each episode, the first-round had finished with the elimination of one of the nine contestants voted the weakest link by their peers. The second-round began, therefore, with just eight contestants. There were two main aims, (a) to test whether the proximity bias that was exposed in the first-round of voting would also carry into second-round voting; (b) to test whether the votes cast in the first-round had an effect on the second-round voting (i.e., reciprocity and consistency) and (c) to test for the effect of consensus.

For (a) and (c) the second-round voting pattern showed the same proximity and consensus effects previously reported in first-round voting of the 72- and 151-episode studies. A cumulative analysis was also provided for proximity and consensus of voting as a function of *voter-votee* spatial relationships using the first and second-round voting from the 72-episode study and first-round voting from the 151-episode study generating an analysis based on 2,583 voting decisions. For (b) second-round voting showed two biases emanating from the first-round of votes. *Voters* in the second-round were significantly more likely to retaliate against the contestants that had picked on them in the first-round. Furthermore, the second-round *voters* that had picked someone in round one were still significantly more likely to target the same *votee* again in the second-round.

**Introduction.** Observations from first-round WL voting showed a peculiar bias for proximity. One argument put forward for studying the first-round voting was the advantage that it was untainted by any previous voting. The contestants voting behaviour in round one is also the least likely to be influenced by strategic voting. The key goal in voting in the first-round was predominantly to return a vote for the weakest link, the “worst player”, based on the TYPE I *game-specific* (i.e., exogenous factor), objective information where available. When this type of information was not available, or it was of low signal strength then contestants picked the weakest contestant using TYPE II, subjective information (i.e., endogenous factor) containing all of the implicit biases that the contestant might harbour. The proximity bias became stronger as the consensus about the weakest link decreased. This was taken to be an indirect measure of low signal strength TYPE 1 information.

This chapter looked at the second-round voting decisions made by contestants in the 72-episode study. One aim was to test whether contestants introduced a third factor into their decision-making by using TYPE III strategic information. This type of information rests on the ability of the contestants to make conscious and deliberate choices to underpin their voting decisions. This kind *strategic-based* discrimination, less influential in the first-round becomes more prominent as the rounds progress (Levitt, 2004; Antonovics et al., 2005). Contestants now have the extra information available to them from a combination of information accumulated from the first-round of questions/answers, the voting decisions from the first-round and the second-round questions/answers to make their second-round voting decision. In particular, the possibility is now open for second-round *voters* to use first-round voting information to explicitly retaliate and reciprocate by

punishing the first-round *voters* that had picked them. This chapter tested for evidence of this kind of *punishment* in second-round voting. Another prediction was that first-round *voters* having already voted a contestant might be consistent by voting the same contestant in the second-round.

Antonovics et al. (2005) found that their evidence for gender discrimination disappeared immediately after the first-round. Another aim for this section was to test whether the proximity bias also persisted into second-round voting.

A key factor that lead Goddard et al. (2011) to attribute the NE was responsible for the voting bias they found, was because the effect was related to the consensus regarding the identity of the weakest link. This relationship should become more pronounced as the consensus regarding the identity of the weakest link decreased.

**Method.** The methods used here have already been described in the General Methods chapter in **M1.2**. This section provides a summary of relevant details pertaining to a reanalysis of Goddard et al. (2011).

***Field observation of WL TV game-show.*** A group of players ( $N = 8$ ) accumulated a pot of money by fielding a second-round of general knowledge questions. Next, each player identified one of their fellows as the “weakest” in that round. The player accruing the majority of votes was summarily eliminated from the show. The second-round voting patterns in 72 episodes of the UK version of WL game-show were analysed. The episodes were originally shown on free to air TV in the UK TV between 24<sup>th</sup> April 2002 and 8<sup>th</sup>

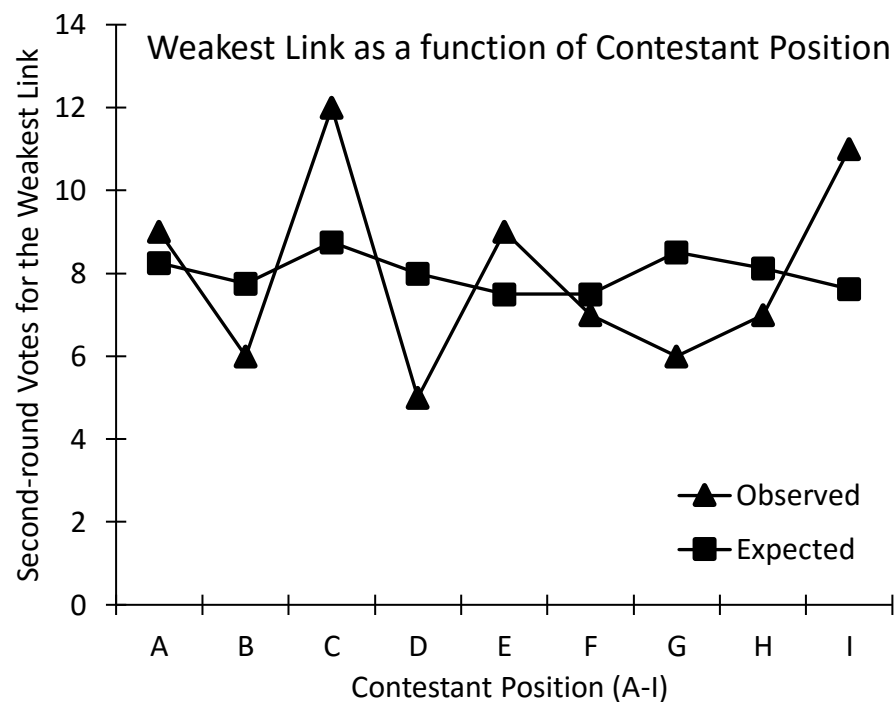
January 2003. Each episode was coded according to the date it was aired using a team of independent coders who were undergraduate students at the University of Lincoln. This part of the study was based on the coding originally undertaken by the independent coders and was not based on observing the original 72 episodes of the WL. Each contestant's position (refer to Figure 3, p. 62) and their gender were noted, as well as their vote. The votes were recorded as integers in the manner previously described in **M1.2**.

**Results.** The data used for this part of the study were the second-round votes from 72 episodes of the WL, each of which included eight contestants, hence a total of 576 (Female = 254, Male = 322) contestants voting activities were recorded altogether. Therefore, each contestant made a seven-alternative-forced-choice (7-AFC) to vote out another contestant as WL.

The analysis used in this section followed the same pattern as before. The *observed* frequency of votes cast by contestants was compared with the frequency of votes *expected* purely due to probability. The first and third parts replicated the analysis of Goddard et al. (2011) regarding voting behaviour and proximity followed by proximity and consensus; the second part tested for any bias from first-round voting affecting second-round voting.

***WL as a function of contestant position.*** The same assumption was made as previously that there should be no systematic bias in the arrangement of the contestants in the second-round of the game-show too. To test this, the position of the weakest link throughout the 72 episodes was recorded and the frequencies that second-round weakest links appeared in positions A - I is shown below in Figure 25. There was no systematic

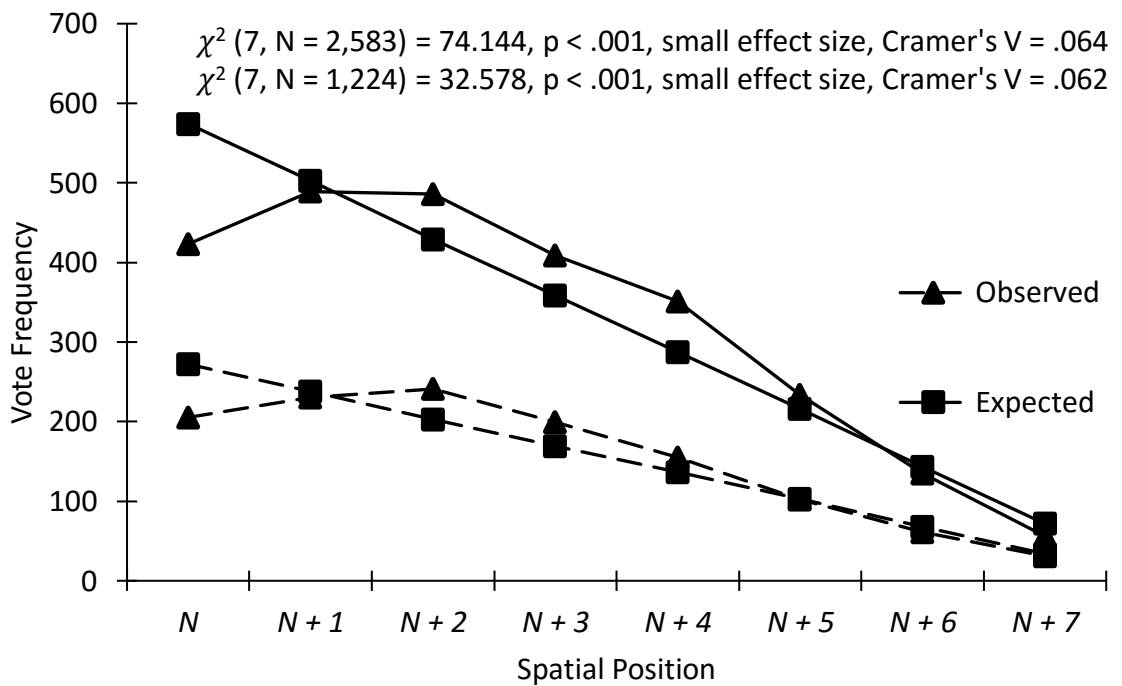
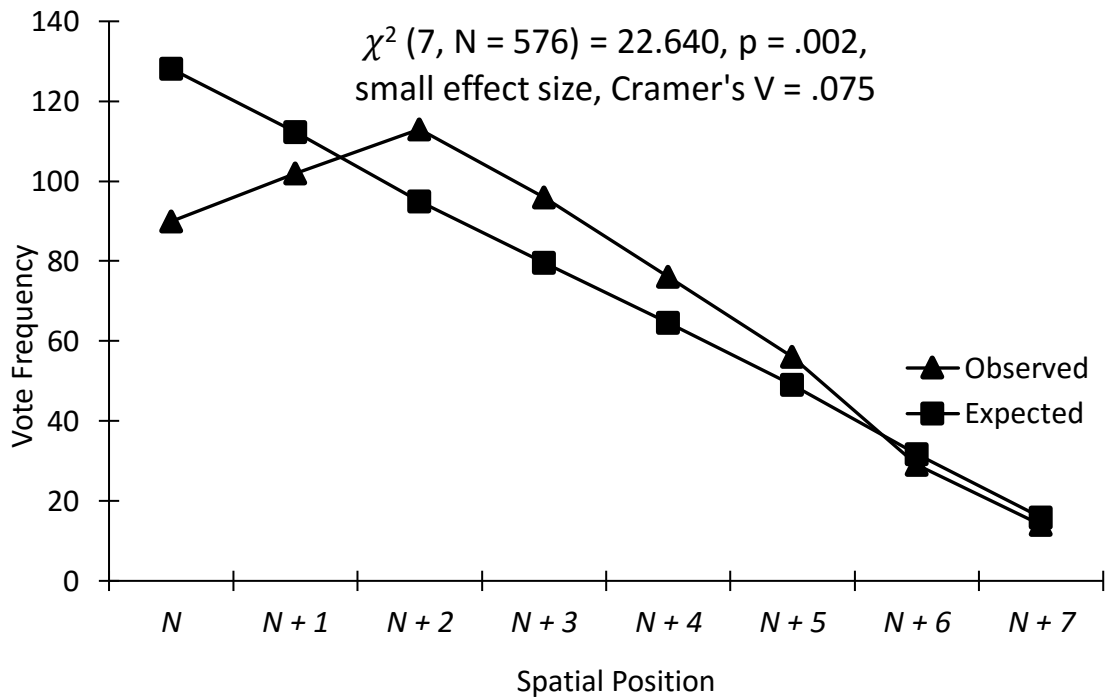
bias in the spatial or temporal order of candidates as the observed frequencies ( $f_o$ ) of the weakest link found in each spatial position (A - I) compared to the expected frequency ( $f_e$ ) was non-significant,  $\chi^2(7, N = 72) = 5.514, p = .702$ , small effect size, Cramer's  $V = .105$ ). Contestants were no more or less likely to vote for candidates that preceded or followed them temporally in the testing sequence.



*Figure 25.* Plots the frequency that the weakest link landed in each of the nine contestant positions (A - I), the expected distribution (square) and the actual are shown (triangle) for the second-round.

***Finding 1: Voter-votee proximity (The “neighbour-avoidance” effect).*** One of the key findings has been the demonstration of a proximity effect in voting. Figure 26.1 (upper panel) shows that this reluctance to vote neighbours as weakest link carries on into second-round voting too.





Figures 26.1 and 26.2. Voting frequency as a function of spatial position. Compares the expected voting frequency (square), derived as the unbiased pattern of voting, with the observed pattern of voting (triangular). Voting frequency is plotted as a function of the

spatial relationship between *voter* and candidate where  $N$  refers to the *voter*'s neighbour and  $N + 1$  the next-door-but-one neighbour and so on. (The data shown here includes the votes cast by the eventual weakest link). Figure 26.1 (upper panel) shows the vote frequencies from the second-round of voting (72-episode study). Figure 26.2 (lower panel) shows the accumulated data from the first and second-round of the 72-episode study (dashed line) and the accumulated proximity data from the 72-episode study with the first-round voting from the 151-episode study (solid line).

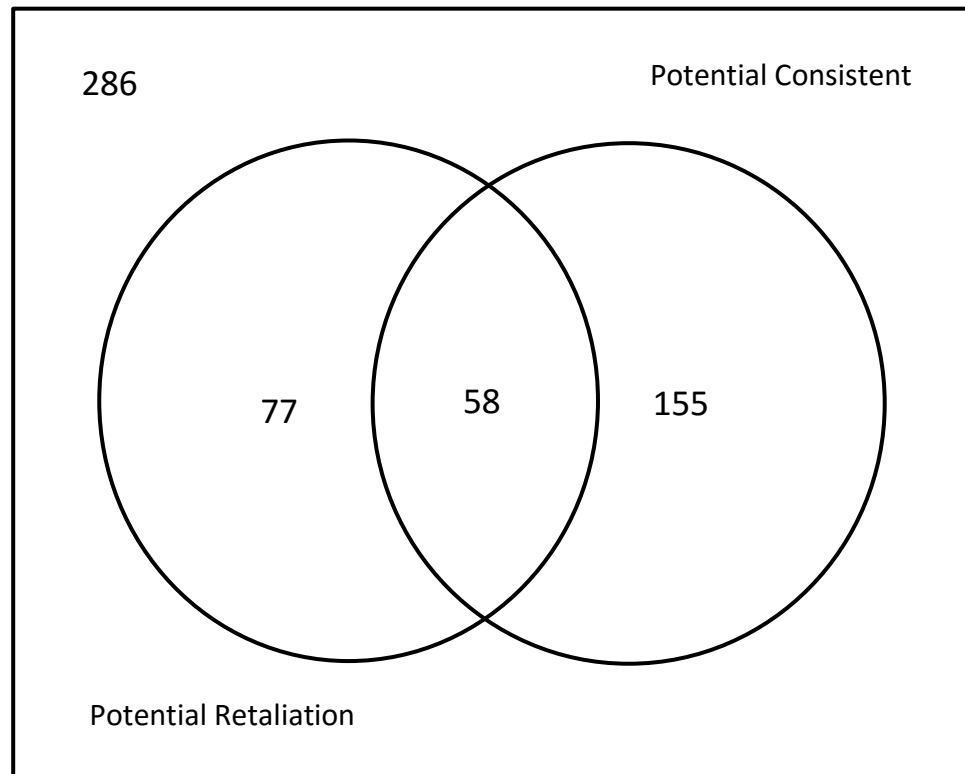
Figure 26.1 shows the expected pattern of voting (square) and observed vote frequencies (triangular) for the contestants following the second-round of questions. The same key features of first-round voting persisted into the second-round voting. The *proximal*/neighbour phase shows that direct-neighbours avoided voting for each other ( $N$ ,  $N+1$ ). The *medial* phase ( $N+2$ ,  $N+3$ ,  $N+4$ ) shows where contestants were most vulnerable to being voted whilst those furthest away in the *distal*/distant phase showed the expected vote frequencies.

Figure 26.2 shows the accumulated vote frequencies from the various studies of WL observed. The dashed line shows the first and second-round votes from the 72-episode study. The solid lines show the sum of the first and second-rounds of the 72 episodes summed with the first-round vote frequencies of the 151-episode study. Taking these observations together reveals a clear effect of proximity on the voting by contestants on the WL TV game-show. What becomes most pronounced is a reluctance for contestants to pick their direct-neighbours as the weakest link.

***Finding 2: The effect of reciprocity/punishment/consistency.*** One bad turn deserves another. The analysis of first-round voting allowed for a consideration of voting untainted by any voting in previous rounds. This part examined whether first-round voting did have any effect on second-round voting. One possibility was that surviving *votees* from the first-round voting experience might seek “revenge” by significantly targeting their first-round *voters*, so reciprocating by voting for them in the second-round. Equally, surviving *voters* from the first-round might be expected to remain consistent in their first-round judgement by targeting the same *votee* again in the second-round. The contestants in the second-round were then identified as belonging to one of the following four categories:

1. Potential *retaliation*: A contestant that survived first-round voting but got voted for by someone anyway; these people were coded as “reactive”, meaning they got voted for in round one and might react by voting for the person who voted for them.
2. Potential *consistent*: Someone who cast a vote for a *votee* who survived to the second-round (i.e., potentially reactive above) was coded as potentially “consistent”; they would be consistent by voting against the same contestant again as in the previous round.
3. Potential *retaliation* and *consistent*: Contestants that fit both the above categories, they got voted for and voted for someone who was not the actual weakest link.
4. *Neutral*: Someone who does not fit into either of the above categories, they did not get voted for nor did they vote for someone who was not the weakest link. These people

were coded as “neutral”. Meaning they voted for the contestant in round one who was the actual weakest link and eliminated from the game-show. These contestants were important by acting as a control group against which to test if the contestants with grounds for *retaliation* and *consistency* actually went on to realise their potential for bias.

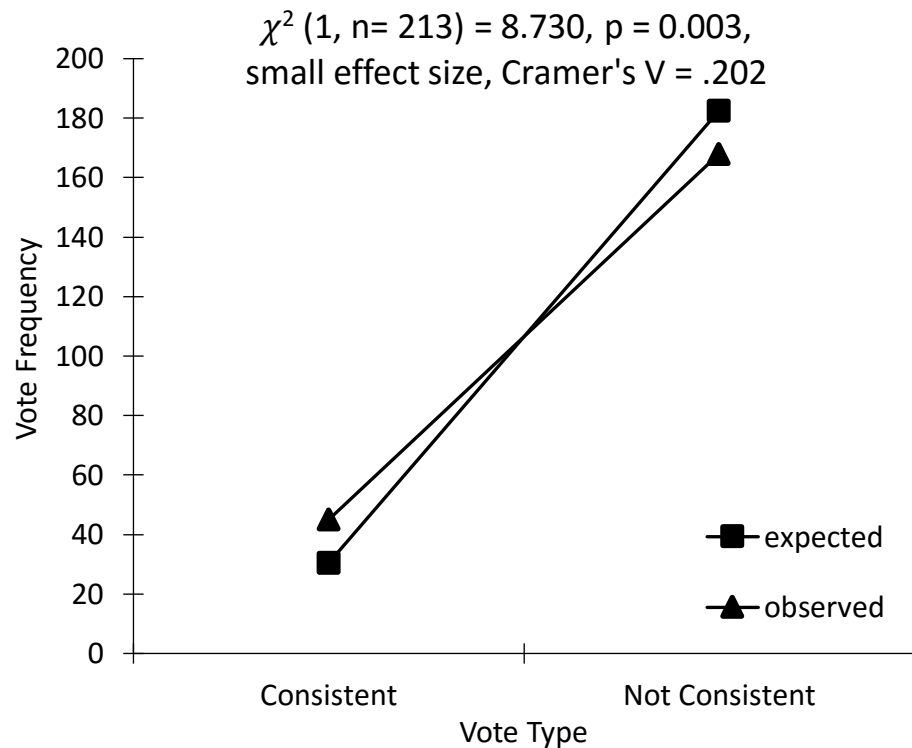
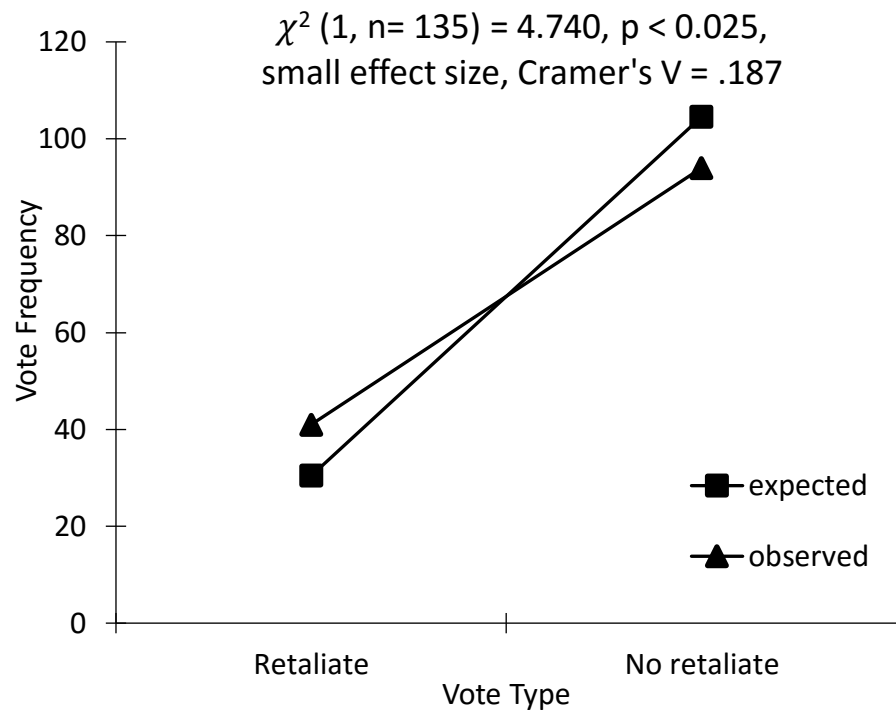


*Figure 27.* Breakdown of contestants in a Venn diagram in four categories. Contestants were grouped according to their potential for *retaliation* and their potential for *consistency* based on whether they were surviving *voters* or *votees* from the first-round. If the potentially vengeful contestants from the first-round exercised tit-for-tat voting, then they would be expected to vote for their first-round tormentors significantly more often than expected.

Figure 27 shows the distribution of contestants by their potential for *retaliation* (*votees* from the first-round) and their potential for being *consistent* (*voters* from the first-

round). From the 576 contestants that made it through to the second-round of the game-show, 286 of them had voted for the first-round weakest link so, they had no one left from the first-round that might seek revenge on them. These 286 contestants also had not received any first-round votes from the other contestant, so, in the context of this analysis they were *neutral* because they had no reason to retaliate against another contestant, at least based on them being voted for. Another 58 contestants, however, were both first-round *votees* and first-round *voters* whose *votee* was still “alive” in the second-round. Therefore, these contestants had the dual potentiality for *retaliation* and *consistency*.

Expected frequencies were again derived in the usual fashion for each contestant in each episode to reflect their potential for *retaliation* and *consistency*. It was fairly easy to work out potential consistency because for those contestants made only a single vote in the first-round, their expected probability to vote for the same *votee* again in round two was always  $\frac{1}{7}$ . The expected frequencies for the contestants that were potentially “vengeful” was a little more complex because they might have had more than one *voter* from the first-round. The computation of the expected frequencies took these factors into account and are shown in Figures 28.1 and 28.2.



Figures 28.1 and 28.2. Shows the second-round voting for contestants that had the potential to *retaliate* (Figure 28.1, the upper panel) and the potential to be *consistent* (Figure 28.2, the lower panel). In both cases the respective categories of contestants show

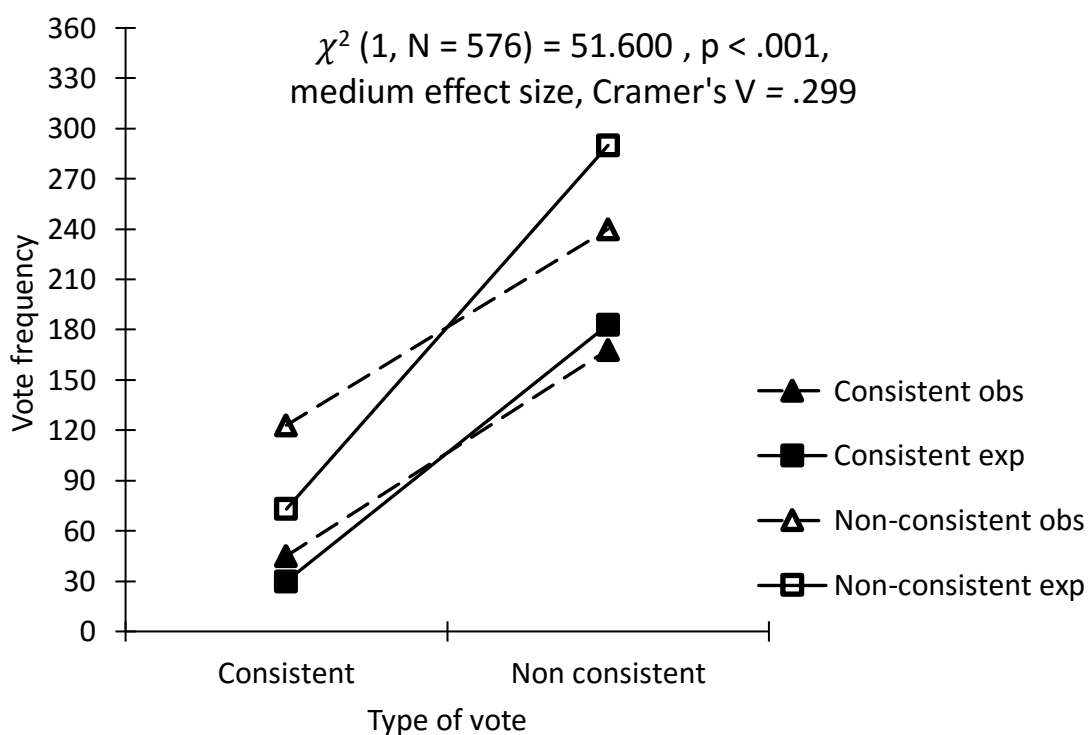
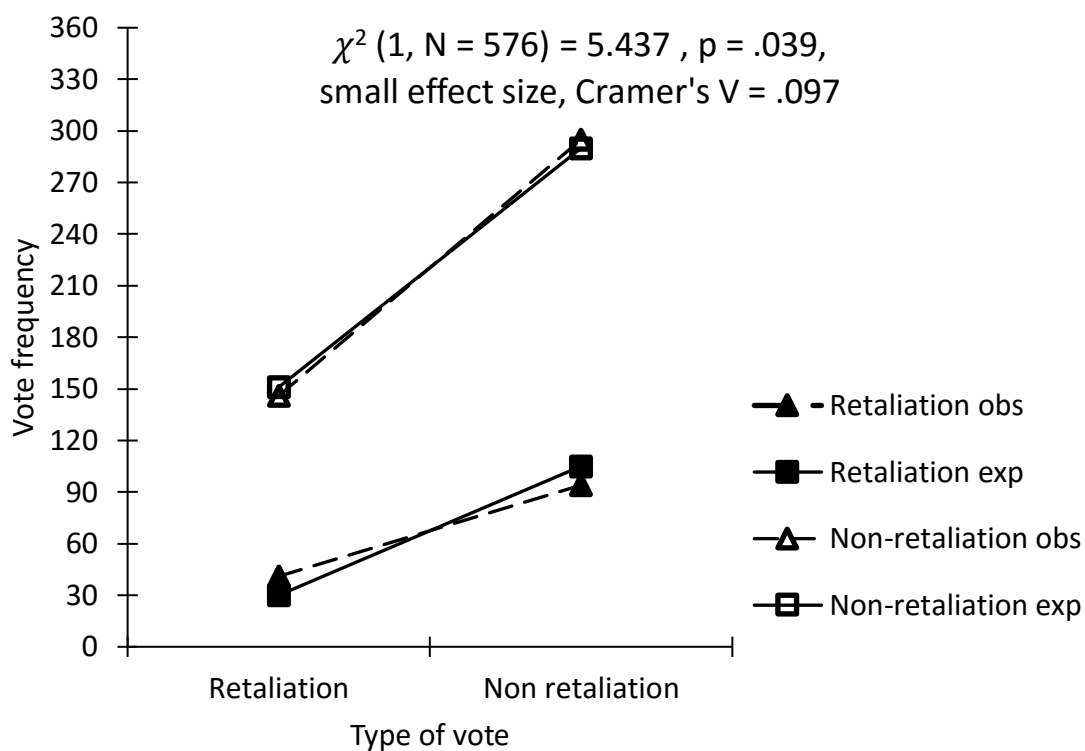
significantly more retaliation and consistency than expected due to chance as indicated by one way  $\chi^2$  tests.

On the face of it, Figures 28.1 and 28.2 show that, the contestants that had been singled out with a vote in first-round voting were also more likely to react by returning the vote in the second-round voting. Also, as predicted (see Figure 28.2), the contestants that identified a *votee* for WL in the first-round were also significantly more likely to be consistent by voting for them again in the second-round.

However, these observations need to be considered in the context of the other voters in the second-round. For example, the potential *retaliators* ( $n = 135$ ) were categorised as such because they were a recipient of first-round vote(s). Although they were more likely to show “retaliation” by returning the vote, there still remains the possibility that they chose their vote based on TYPE 1 *game-specific* information rather than utilising TYPE III *strategic* information in the form of “cold revenge”. In order to test whether their votes were motivated by “revenge” then the *retaliators* choice must also be compared with the votes made against the same *votees* by the other contestants in the second-round that had no grudge or reason for retaliation against the *votees*. This data is shown in Figure 29.1 (upper panel). Those contestants ( $n = 441$ ) that had no reason for retaliation would nevertheless still be expected to vote for some of them through chance ( $n = 151$ ). The potential retaliators voted according to retaliation ( $\frac{41}{30} = 1.367$ ) compared to their non-retaliatory controls ( $\frac{146}{151} = .967$ ) that performed almost at chance levels. This suggests that the *retaliators* were motivated by revenge! ( $\chi^2(1, N = 576) = 5.437, p = .039$ ; small effect size, Cramer’s  $V = 0.097$ )

Figure 29.2 (lower panel), tells a rather different story regarding the consistency of contestant votes. Again, on first analysis it seemed that the contestants that had the potential to vote consistently, did so by being significantly more likely to vote for the same contestant again in the second-round. However, once the control was introduced, it revealed that the contestants with no grounds to be consistent still targeted those same contestants significantly more than expected. This suggests that there are some other factors, other than consistency to target these contestants. In this case, it was more likely that voting was based on TYPE 1 *game-specific* information than TYPE III *strategic* voting. In fact, the consistent *voters* show a reluctance to vote for the contestants than the neutral controls, suggesting that their targeting of them in the first-round might make them more reluctant to vote for them again. The proportion of votes for the consistent *votee* from the potentially consistent second-round *voters* was 1.5 compared to 1.68 by the controls ( $\frac{45}{30}$  and  $\frac{123}{73}$ ).





Figures 29.1 and 29.2. Show the same data as in Figures 28.1 and 28.2 with second-round voting for contestants that had the potential to be purely retaliatory (Figure 29.1, the upper panel) and the potential to be consistent (Figure 29.2, the lower panel). This time their

performance is compared with their respective counterparts that had no potential for retaliation and consistency.

***Finding 3: The effect consensus.*** Consensus represents an important independent variable because of the underlying assumptions. When there was high consensus, for example where all eight contestants (i.e., aside from the weakest link themselves) chose the same *votee* then there was some degree of certainty over the identity of the weakest link in that episode. Likewise, low consensus rounds were indicative of uncertainty. Uncertainty is the key factor as to why contestants will show prejudice, discrimination, proximity biases or any other kind of bias in making their voting decisions. Therefore, the proximity bias should one exist, should become more pronounced under the conditions of greatest uncertainty. The assumption is made that contestant uncertainty will be related to consensus so that the neighbour-avoidance effect should increase with a decrease in consensus. In order to investigate this, the episodes from the 72-episode study (first and second-round, 144 episodes) and the 151-episode study were considered together. In doing that a better distribution of episodes across the levels of consensus is generated ( $N = 295$ ). Table 12 shows the distribution of consensus across the 295 episodes and plotted in Figure 30. Notice that in the actual episodes observed, voting tended towards the higher consensus over the WL ( $\chi^2(2, N = 295) = 1504.398, p < .001$ ; large effect size, Cramer's  $V = 1.597$ ).

Table 12  
The Contestants Consensus Votes in 295 Episodes

Consensus	295-episode	Null
1	0	0.293413699
2	14	134.3925118
3	51	130.5619043
4	51	26.48635805
5	56	3.039672375
6	45	0.217119455
7	46	0.008862019
8	32	0.00015825
Total	295	295

*Note.* The table shows the frequency distribution of episodes (N = 295) by consensus, the number of votes cast for the eventual weakest link; first-round and second-round voting of 72-episode study and first-round voting of 151-episode study.

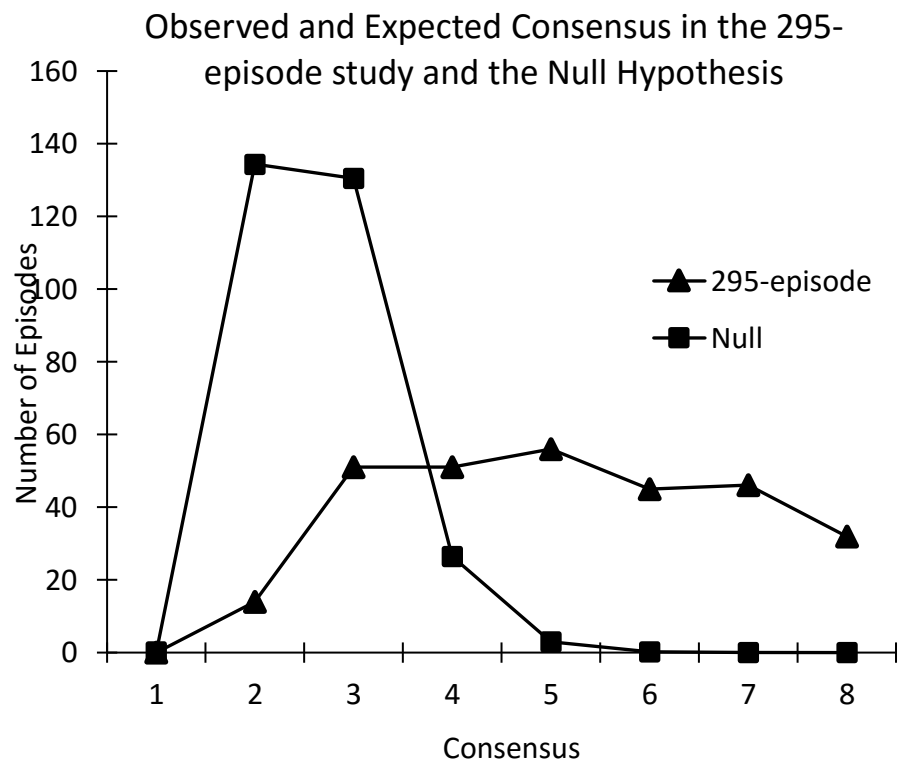
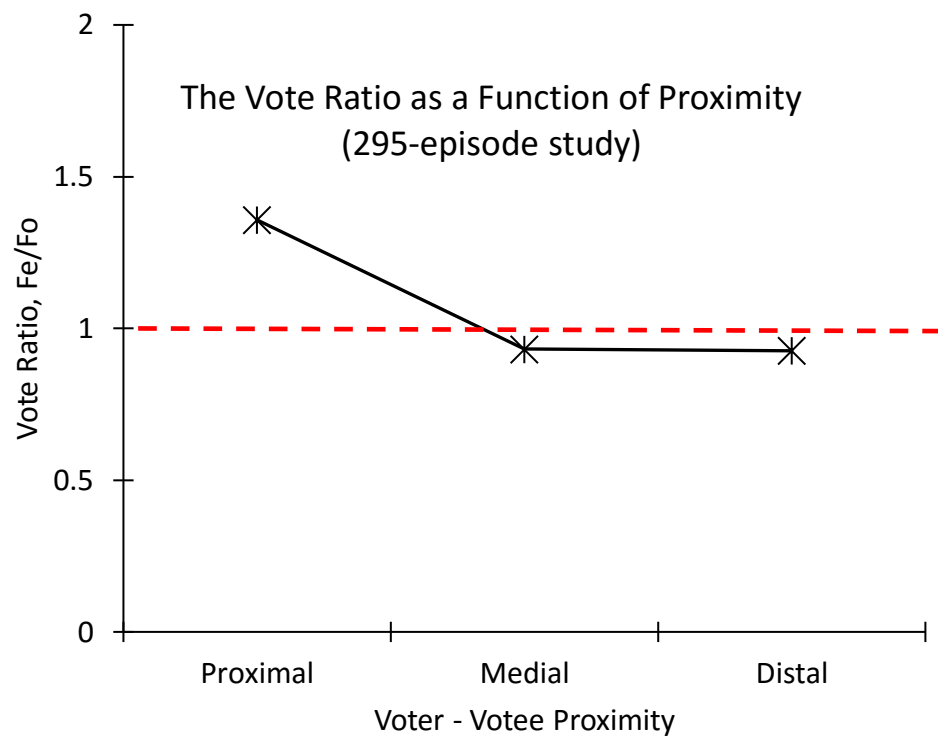


Figure 30. shows The comparison between the observed accumulated first and second-round voting of 72-episode and 151-episode studies with the null hypothesis distribution of votes.

*Partitioning Consensus.* Figure 31 below shows the vote ratio as a function of proximity for 295-episode study. A similar pattern was observed in 72-episode and 151-episode studies where in all three studies, it displays a positive bias at *proximal* proximity, 1.357 where the ratio value was more than one. Similarly, in medial and distal proximity, the ratio values were less than one which suggests either no bias or a slight negative bias towards participants who were medial or distal.



*Figure 31.* shows the vote frequency ratio as a function of spatial proximity for the 295 – episode study. The proximity is divided into three phases (*proximal*, *medial*, and *distal*). The red line indicates the ratio is equal to one as the baseline.

**Discussion.** Three main findings emerged; (a) contestants still avoided voting their direct-neighbour (s) as the weakest link, (b) surviving *votees* from the first-round were significantly more likely to “punish” the *voter* in the second-round, (c) the *voters* that elected a *votee* in the first-round were still significantly more likely to be consistent and vote for them again in the second-round. Other features from the second-round voting are discussed in the General Discussion.

## R1.5 General Discussion

The three studies presented in this chapter all used the same pseudo-naturalistic field observation of the WL game-show. The unique features of the game-show underlined an implicit bias that exists in decision-making and proximity. Another bias was discovered regarding the *voter-votee* gender. Finally, a third kind of bias was demonstrated based on reprisal and consistency with first-round voting.

**Proximity and consensus.** The main finding was the demonstration and repeated replication of the proximity bias in voting. It was most clearly manifested in three phases, where the *proximal* phase was characterized by the *neighbour effect* (NE): a significant reluctance to do something negative to close neighbours. A *medial* phase where intermediate contestants were more likely to be voted against and a *distal* phase where the most distant contestants received the expected vote frequency.

The *proximal* bias occurred as a NE and it is suggested the *distal* phase occurred because the voters were reluctant being in a situation where they had to mete out “punishment” towards the recipient. Such behaviour can even be construed in a sociological perspective as an interaction order bias (Goffman, 1983) whereby the physical proximity affects the doctrine to punish an individual who is closer to the punisher, such as in the case of the *voter*.

It is proposed that, the observed NE happened when the signal strength of TYPE I information was low. Trope and Liberman (2010) suggested that this occurrence can be

explained with construal level theory and psychological distance. The greater the *voter-votee* distance becomes then the TYPE II information is less visibility, therefore the *voter* is assumed to have low construal level towards contestants that were physically and socially distant. By contrast, contestants that are close in distance, or high construal level, such as direct-spatial neighbours, were considered to have a concrete visibility for the *voter* to make a more rigorous judgement.

However, information availability is an important element for the contestants to cast their vote. NE was less profound in situation with high explicit information availability, for example, in episodes where a single contestant constantly gave incorrect answers and was easily identified as the weakest contestant. Thus, in the scenario of high voting consensus, it is suggested that the decision made relied on the observed cue during the game session. The public, situational factors accounted for the *voter's* choice. This was made using the conscious, slow system 2 decision-making.

Conversely, when the situational information signal strength was low, or its availability was low, for example, when all the contestants were performing well with correct answers: the *voters* had a tough decision to make. One compromise choice, to avoid making an uncomfortable atmosphere for the next round, would be to avoid picking on neighbours. This would appear a sensible strategy especially given that contestants likelihood for tit-for-tat in the second-round voting shown in the **R1.4**. Hence, it is suggested that the decision made with low information TYPE I availability, used dispositional TYPE II information, processed in a different stream of the dual process theory. The ability to process the information due to limited external cues, meant that the

contestants made a fast and automatic decision, yet without conscious reasoning. The *voter* used their System 1 to make their decision followed by System 2 to justify when asked to.

The most compelling aspect of the results relates to the effect of consensus where in all of the examples the strength of the proximity bias increased with reduced consensus. This was most pronounced when the many episodes from all of the observations were concatenated in **R1.4**. This allowed for sufficient observations to partition the episodes according to consensus. Under these conditions there was a general trend for contestants to become increasingly more likely to avoid neighbours, at least up to a consensus of four. Lower consensus episodes (2 and 3 consensus) still had a pronounced NE but less than consensus four. It is uncertain what the cause of this is or whether it has any deeper significance.

**Gender.** The first study (see **R1.2**) carried out a reanalysis of Goddard et al. (2011) 72-episode study confirmed their findings on proximity and consensus but also revealed a surprising gender effect. Goddard et al. referred to a gender effect but this was mainly restricted to the disproportionately unequal numbers of male and female first-round weakest links. A closer examination of their data, however, exposed a much more systematic gender effect as females were much more likely to be voted against, both by males and other female contestants. This surprising finding was consequential in two respects. Firstly, it went against the previous findings on discrimination where females and males were more likely to vote for females not against them (Antonovics et al., 2005).



Secondly, and especially pertinent to this thesis, was the possibility that the proximity effect was merely an artefact of the gender voting bias.

The reason for suspecting the latter was that as an observational field study there is no control over randomisation of participants to conditions. A tally of the contestant configurations found a bias with nearly half taking the same arrangement (i.e., Male-Female-Male-Female-Male-Female-Male-Female-Male). Given this dominating pattern allied with the gender effect it is not surprising if females were to avoid neighbours as they were mostly men. It turned out though that breaking proximity voting down by gender still revealed a strong NE. Moreover, as the consensus reduced the NE increased as predicted. Nevertheless, the gender effect and proximity effect required further scrutiny. This was particularly warranted because the data coded by Goddard et al. (2011) was only restricted to voting behaviour and gender with no game-specific noted (e.g., accuracy of questions answered; banking et cetera). This means that it is not possible to ascertain whether the gender bias occurred as some taste-based or information-based discrimination. The two possibilities existed that, the contestants were prejudiced and harboured deep rooted negative stereotypes about women and their abilities on game-shows, taste-based discrimination. Another possibility was that the discrimination was information-based due to the poorer performance of women in the first-round. As the reanalysis was based on the coding of the episodes rather than the actual episodes themselves then this could not be resolved. Recall that Antonovics et al. (2005) showed that women discriminated against men even though there was no basis for this at least according to their ability to answer questions accurately.

In light of these new revelations in the reanalysis of Goddard et al. (2011) 72-episode data, a replication was needed. This prompted perhaps the major work of the thesis in the form of the 151-episode replication. The findings that came out of the 151-episode study once again confirmed the proximity bias but also displayed a different gender bias more in line with Antonovics et al. (2005). It showed that there was a bias for women not against women and that this bias was strongest for female *voters*.

**Limitations.** The demonstration of NE occurred in a quasi-naturalistic setting and the advantage of this study is that, it is free from demand characteristics where the observed behaviour was performed without any interference from the researcher. The contestants were taking part in a real event where they believed that, they were taking part in a TV game-show, such that the votes cast were derived from the genuine setting and rules of the games.

However, the limitations inevitably arise from this field observation. Principal amongst them being, is first, lack of control on the salient variables (e.g., gender), second, the selection of contestants (e.g., demographic factors), third, non-random allocation (i.e., at least five different contestant's arrangement by gender) and finally, the contrived setting of the TV studio.

It remains to be seen whether the NE was observed as a peculiar artefact of the TV studio. Moreover, it remains uncertain as to whether the NE would be evident in other social scenarios. Further, the vote in the context of the weakest link was construed as being negative, nasty or bad in some way. It also remains uncertain as to whether the NE

would still occur if the contestants were making a vote of approval, by for example, picking the strongest link. The NE could just be a consequence of picking from a one - dimensional (1D) array, irrespective of the value of the choice being made. These questions however, cannot be addressed by recourse to an observational field study. The only way that they can be resolved is by designing scenarios that come under some kind of experimental control. The next two results chapters record the attempts to demonstrate the NE outside the confines of the WL TV game-show.

## Chapter R2

### Testing the Neighbour Effect using Voting Activity

#### R2.1 Chapter Outline

**R1** demonstrated a compelling *neighbour effect* (NE) that emerged as a bias in the voting behaviour of contestants on the WL. For this demonstration to have any wider significance though it also needs to be exhibited in contexts other than the TV studio. The aim of this section was to test whether the conditions could be designed to replicate the NE. This section presents investigation into the NE in the different social context of a standard lecture hall populated by first-year undergraduates. The participants were presented with a kind of voting task to pick another person in the lecture theatre to receive a reward or punishment, by adding or removing their chances to win a prize draw.

The rest of the chapter is in four sections: a general introduction focuses on the effect of reward and punishment to understand the NE, the methods give the details of the experimental set-up, results and discussion.

R2.1 Chapter outline

R2.2 Introduction

R2.3 Methods

R2.4 Results

R2.3 Discussion

**Abstract.** R1 demonstrated how individuals (i.e., contestants of the WL TV game-show), avoiding doing nasty things towards their direct-spatial neighbour(s), which in this case meant voting out another contestant as the weakest link. This behaviour was denoted as the *neighbour effect* (NE). At this point, it seems that the NE might be a profound bias that affects an individual when performing *nasty* or negative actions against another individual who is closest to them by physical proximity.

One issue that remains unresolved concerns whether the NE occurs merely as an effect of voting. Hence, one test would be as to whether the NE also occurs if tested in a way where an individual ought to be *nice*. Participants were the first-year students (N=449) attending an induction lecture during their orientation week. They were asked to cast a closed, secret vote, for another person in the (a) same row, (b) same block and (c) the lecture theatre by marking a 'X' on a seating plan. The vote carried either a positive or negative outcome for its recipient (i.e., another participant) by adding or removing their chances to win a prize draw. The findings showed that the participants who cast negative votes demonstrated a significant NE by avoiding voting for their nearest neighbour(s). However, a reverse pattern was found when participants gave positive votes.

## R2.2 Introduction

The empirical demonstration (refer to **R1, p. 77**) of the *neighbour effect* (NE) was observed from an observational field experiment of the WL TV game-show. The study was replicated suggesting that the NE was a robust bias in WL voting. However, despite the strength of the demonstration of the NE, field observations, although conferring special advantages, fundamentally lack the required control to scrutinise the phenomenon further. This leads to the following predictions and tests regarding the robustness of the NE and its nature.

1.     **Replication.** The demonstration of the NE in the WL should be tested for its robustness by being replicated in other voting scenarios and
  
2.     **Valence.** Given (a) if the NE occurs as a simple spatial bias in voting per se, then the NE would be expected irrespective of vote valence. That is, if there is an inherent avoidance of selecting neighbour(s), then it will occur whether the consequences of choosing them are positive/ neutral/negative. If, however, the NE is selective for vote valence, by only occurring when the vote has a negative valence, then it shows that it can be interpreted as a social preference. Given that people tend to avoid conflict more when an outcome is potentially positive (Weber & Camerer, 1998) then the NE might also be expected to switch if the valence of the vote (i.e., positive or negative) changed.

**Does Prize Size Matter.** Consider the unique context of the WL as experienced by the contestants. The NE was observed in the rarefied atmosphere of the TV studio, where contestants were aware of playing the game with strangers being faced by an intimidating host, whilst being filmed for national TV and facing a TV studio audience. Any of these factors alone would be sufficient to raise the stress and adrenaline of the contestants but taken together produced a highly charged incentive to play the game well. Further incentive, should they be needed, included the value of the cash prize on offer and the desire not to appear foolish on TV. Questions have to be raised therefore about the NE and whether it is just something that occurs in the rather peculiar surroundings of a TV recording studio. It could be argued that the observed behaviour exists in an environment where the conditions such as the game rules were well defined and the prize values were so high.

Several studies have shown though that an increase in the prize size does not necessarily significantly affect the behaviour of decision-makers. For instance, Kocher, Martinsson and Visser, (2008) claim that an increase in the prize values did not significantly affect behaviour in their study (e.g., cooperation). They investigated the effect of stake size in a form of public goods game in a field experiment with a standard laboratory setting. The experiment was conducted with 120 high school participants in South Africa. The experiment consisted two parts, in which, the first part of the experiment was to ask the participants to indicate their preferred contribution to the groups without fear of receiving any punishment, while, the second part was to suggest their contribution would have some consequence or punishment. The participants were divided into two groups, one with low stakes size (LOW) and the other group with high stakes size

(HIGH). Participants were randomly assigned into two groups and they were briefed about the task and been reminded not to communicate among participants to avoid unnecessary information leakage. Since the participants were attending the same school and the chances were higher for them to share information because of familiarity in that particular context, Kocher et al., scheduled the session to be executed at the same time. Their findings showed that, in both treatment LOW and HIGH, there were small differences in both tasks. For contribution task without punishment, the average contribution level in LOW is 34.4% while, in HIGH is 32.9%. In contribution task with punishment, both group given 41.2% in LOW and 40.9% in HIGH. However, for punishment level, their study found that, the participants directed their punishment towards those who were either contributed less or more than themselves. Hence, their study suggested that the increase in the stake size had no significant effect either on their level of contribution or on the punishment. However, a potential drawback of this study was that they were conducted among high school students, who might not understand the consequences of contribution and punishment despite being briefed and tested if they had understood the task.

The findings were consistent with a study done by Herrmann, Thöni and Gächter (2008) in their large scale study among 1,120 University undergraduate students in public goods games. They also found that, individuals punished those who consistently contribute lower but at the same time, they also punished those who contribute more. In which, this statement is true in the sense of fairness where in whatever situation, “People might punish not only freeloaders, but cooperators too” (Herrmann et al., 2008, p. 1,362).



In comparison, in a study conducted by Fehr, Tougareva and Fischbacher (2014) on gift exchange experiments, found that the stake level was less important than the situation itself. Their study focused on the reciprocal fairness, in which the experiment was conducted among 120 first and second year undergraduate from a college in Russia. The participants were divided into two groups whereby each group were split into normal and high stakes. Despite the difference between stakes there were no significant effects, with relatively minor differences in their effort levels.

Parke and Parke (2013) stress the importance of real life scenarios by, for example, using data from the real gambling environments to understand the effect of the prizes and stakes. Stake or prize value is regarded as important element especially in gambling, where individuals willing to take risks with assumption that in return the higher the risk, the better they gain. They suggested that, there were massive limitation in laboratory experiments particularly to test and understand the value of the stakes, which was not possible for any experimenters to conduct a huge scale experiments in comparison to casino. In contrast, the NE displayed in **R1** was first observed in a TV game-show, where the prize is considered huge in comparison to other methods. Therefore, in this study, it is important to test the NE in a more controlled environment so as to assess if NE is a robust bias in decision-making.

***Aim, Objectives and Hypotheses.*** The aim of this study was to test the NE in a different setting (i.e., non-exclusive environment as a TV game-show) and to understand the effect of vote valence in NE. Hence, to achieve the aims of this study, the objectives of this study were;

1. To test whether the NE bias in decision-making is a robust endogenous factor.
2. To test the effect of the exogenous factor, vote valence, on NE.

Thus, to test whether the NE, and spatial proximity, occur in contexts other than the TV game-show. the hypotheses were;

1. The null hypothesis, as in the observations of the WL, was the expected pattern of voting due to chance.
2. Based on the strong NE on the observation of WL then a similar neighbour effect was predicted here.
3. Another test concerned identifying whether vote valence was an endogenous factor in decision making and follows on from whether an NE is found in 2.. If valence has no effect on NE then it would imply that it is not a factor at all. A universal NE lacks sensitivity and is of limited interest but a change in voting according to valence shows that there exists a sensitive and powerful endogenous bias working at an implicit level.

## R2.3 Method

**Participants.** Participants were undergraduate *freshers* attending their first orientation lecture at the University of Lincoln (N = 449; Female = 340 (75.724%), Male = 85 (18.931%), Undisclosed = 24 (5.345%), see Table 13). The participants' responses were collected from two different cohorts (2013/2014) and (2015/2016).

Table 13

*The Participants' Gender (n=449) for 2013/2014 and 2015/2016 Cohorts*

Characteristic	n	%
Gender		
Female	340	75.724%
Male	85	18.931%
Undisclosed	24	5.345%

**Procedure.** For this study, the participants were asked to take part in a *Voting* activity. The *Voting* task took approximately 10 to 15 minutes to complete which included introduction and debrief. The procedure was explained in the Methods Chapter, **M1.4**. For this activity, the main aim of this task was to get the participants to cast a private vote that carried either a positive, neutral or negative (i.e., -5, -1, 0, +1 and +5) effect for its recipient. Five levels of valence (i.e., +1, +5, 0, -1 and -5) were used in the 2013/2014 study, but, the 2015/2016 study only used the extreme valence levels (i.e., +5 and -5). The positive vote (i.e., +1 or +5) added to the chances of its recipient winning a raffle by increasing the number of virtual 'tickets' the participant received. The negative vote (i.e., -1 and -5) reduced the chances of the recipient winning the raffle by the amount stated. The neutral vote (i.e., 0) carried no weight and its recipient was neither advantaged nor

disadvantaged by its receipt. Irrespective of the total tickets accrued by each participant during the voting activity it was ensured that every participant received at least a minimum of one 'ticket' for entry into the raffle. All of the tickets were virtual. Subsequent prizes were course related text books for the (2013/2014 cohort) or Amazon Gift Vouchers (for the 2015/2016 cohort). Three winners were picked randomly for each cohort following the orientation lecture.

Each participant also received a *credit point* (CP) that allowed them to open an account with School of Psychology Credit Point System, this served as a means by which to encourage undergraduates to volunteer as participants within the research community of the School of Psychology. The principal aim of the orientation lecture was to introduce the new students to the School of Psychology tutorial system and Credit Point Scheme. Participants were instructed that the main aim of the lecture activity was to encourage them to register their CP online account to redeem their CP. Participants were reassured that they would receive the CP whether they participated in the task or not and would be entered into the raffle also. Although signed informed consent was not sought participants were assured of their anonymity and no identifiable personal details were recorded. Participants were also reminded of their right to withdraw and were told that if they did not wish to carry out the vote then they could simply submit their ballot paper blank in the envelope provided. If they subsequently decided to withdraw their data, then they could do so easily by making a note of their seat number and contacting the researchers. It was stressed that the activity was private and that their choice would not be revealed.

Four different versions of response sheets (refer to Appendix A1-A10) were randomly assigned to the participants to ensure the reliability of the data collected. The response sheet served as their *ballot* paper, which included the seating plan for the lecture theatre (refer to Figure 5 for the seating plan). Each participant was required to vote for a candidate that was another participant, (a) sitting in the same block and in the same row of seats, (b) sitting in the same block and (c) sitting anywhere in the lecture theatre.

**Observation from the voting activity.** The voting activity data was collected from two different batches of the first-year undergraduate Psychology students by using the same procedure. However, the changes were made at the sequence of the questions and randomly distributed with different versions. The participants seat numbers (i.e., participant's position in the lecture theatre) and gender were noted. Each position was denoted by a unique seat number with alphabets according to each row (i.e., total number of observed row was 21) in the lecture theatre layout, with the first row for all three blocks were left empty (i.e., with the exception for disabled students). Table 14 shows the differences in row designated with different number of seating plan for two different batches. Hence, the possible maximum range for the voting distance were from nine-spaces away ( $N+8$ ) to 15-spaces away ( $N+14$ ) between the *voter* and the *recipient*. However, for this study, the maximum expected voting distance between *voter* and the *recipient* was corrected to 14-spaces away ( $N+13$ ).

Table 14

*The Observed Seats Available for Each Row by Block in the Lecture Theatre*

Row by block	2013/2014	2015/2016
	Seats	
A1	12	12
A2	16	16
A3	11	11
B1	12	12
B2	15	15
B3	12	12
C1	12	12
C2	14	14
C3	12	12
D1	12	12
D2	13	13
D3	13	13
E1	12	12
E2	12	12
E3	13	13
F1	12	12
F2	11	11
F3	14	14
G1	10	12*
G2	10	10
G3	14	14

**Note.** The letters represented the row and the blocks were represented by numbers (e.g. A = Row A, 1 = Block 1). \*The additional seats added in the lecture theatre at G1 for (2015/2016) batch.

***Expected voting pattern based on 14-AFC for participant in the same block in the same row.*** The expected voting pattern was calculated by replicating the similar procedure as discussed in **R1**, by using rudimentary probability theory, in which, assuming each of the votes by the *voter* was always one over the available *recipient's* distance from the *voter*. For all the expected voting frequencies, its calculated by each row by block for each batch based on observed participants seating arrangement and voting valence (i.e., -5, -1, 0, +1 and +5).

The expected frequencies were computed from each single position and added up to complete the overall expected frequencies for 449 participants to derive. Table 15 shows the expected frequencies calculated separately by each batch and each valence.

Table 15

*The Expected Frequencies for the Voting Activity by Cohort*

	Cohort	N	N + 1	N + 2	N + 3	N + 4	N + 5	N + 6	N + 7	N + 8	N + 9	N + 10	N + 11	N + 12	N + 13	Total
Positive	2013/2014	15.662	14.061	12.426	11.091	9.372	7.681	5.892	4.497	3.255	2.064	1.319	0.271	0.205	0.138	87.933
	2015/2016	18.850	16.875	14.484	12.181	10.088	8.536	6.738	5.334	4.108	2.697	1.639	0.605	0.225	0.071	102.433
	<b>Total</b>	<b>34.512</b>	<b>30.937</b>	<b>26.910</b>	<b>23.272</b>	<b>19.460</b>	<b>16.217</b>	<b>12.630</b>	<b>9.831</b>	<b>7.363</b>	<b>4.761</b>	<b>2.958</b>	<b>0.877</b>	<b>0.430</b>	<b>0.210</b>	<b>190.367</b>
Negative	2013/2014	17.507	15.751	13.987	12.267	10.433	8.580	6.965	4.963	3.271	2.013	0.963	0.552	0.410	0.271	97.933
	2015/2016	20.816	18.636	16.339	13.954	12.025	9.306	7.282	5.398	3.871	2.530	1.280	0.605	0.368	0.071	112.483
	<b>Total</b>	<b>38.323</b>	<b>34.388</b>	<b>30.326</b>	<b>26.221</b>	<b>22.458</b>	<b>17.886</b>	<b>14.247</b>	<b>10.362</b>	<b>7.142</b>	<b>4.544</b>	<b>2.243</b>	<b>1.158</b>	<b>0.778</b>	<b>0.343</b>	<b>210.417</b>
Neutral	2013/2014	8.831	7.920	7.052	5.838	5.124	4.399	3.535	2.664	1.617	0.582	0.229	0.138	0.071	0.000	48.000
	<b>Total</b>	<b>8.831</b>	<b>7.920</b>	<b>7.052</b>	<b>5.838</b>	<b>5.124</b>	<b>4.399</b>	<b>3.535</b>	<b>2.664</b>	<b>1.617</b>	<b>0.582</b>	<b>0.229</b>	<b>0.138</b>	<b>0.071</b>	<b>0.000</b>	<b>48.000</b>
	<b>Grand total</b>	<b>81.667</b>	<b>73.244</b>	<b>64.288</b>	<b>55.331</b>	<b>47.042</b>	<b>38.502</b>	<b>30.413</b>	<b>22.856</b>	<b>16.121</b>	<b>9.887</b>	<b>5.430</b>	<b>2.173</b>	<b>1.279</b>	<b>0.552</b>	<b>448.783</b>

*Note.* The positive and negative voting are the accumulation of +1, +5, 0, -1 and -5 respectively. The overall total Exp. (expected) frequencies were used to plot the expected voting pattern.



## R2.4 Results

The lecture based voting activity was designed to test if the NE occurred in a setting different to the WL game-show. The results from the voting activity are presented in two parts to test for the effect of proximity and the effect of vote valence. The effect of gender was not tested for and no predictions were made regarding gender. The opportunity sampling procedure used meant that the distribution of gender within the study was heavily weighted towards female participants (female participants were 75.724% from the overall combined sample 2013/2014 and 2015/2016 cohorts). Given the distorted imbalance in the uneven distribution of gender in the lecture based studies; the failure to find significant gender biases in the 151-episode replication study of WL; and differences in seating configurations between the Lecture based and the 151-episode study; gender was not analysed as a main factor in the lecture based study. Descriptive analysis of gender based voting suggested that gender had no impact on voting.

*The effect of proximity.* The WL TV game-show demonstrated that NE was a significant effect of direct-neighbour(s) avoidance when the contestants were required to perform negative voting (Goddard et. al., 2011; Goddard, 2012; Goddard, Hylton, Parke & Noh, 2013). For this study some of the key features of the WL TV game-show format were simulated to test for the NE in a different setting to the TV game-show. **R1** concluded that the unique conditions of the WL that give rise to the proximity based NE were:

1. Negative vote: The negative vote of WL was simulated by including a negative valence voting condition.

2. **Uncertainty:** The key factor that was identified as contributing to the NE was the *voter's* uncertainty regarding the identity of the WL. When consensus was low regarding the identity of WL the NE proximity bias increased. This study was designed so that there was maximum uncertainty about whom to vote for. At least in the WL there was some TYPE I *game-specific* information or even TYPE III *strategic-based* information. However, in the lecture study there was no information or guidance of any kind regarding how to choose to allocate the vote.
3. *Voter-votee* spatial relationship was clearly defined: Both the WL and lecture study had a clearly defined spatial relationship that could be easily measured and recorded.

Table 16 shows the calculated  $\chi^2$  for each *voter-votee* spatial relationship in the lecture theatre. The analysis from the votes cast by each participant shows that (2013/2014 and 2015/2016 cohorts), there was a significant association between the vote cast and the *voter-votee* spatial position,  $\chi^2(13, N=449) = 69.454, p < .001$ , small effect size, Cramer's  $V = .109$ .

In Figure 32 below, the observed (triangular) and the expected (square) votes were plotted according to the spatial relationship of *voter-votee*. The analysis shows that, a similar pattern emerged in this study in comparison to the study from the WL (refer to **R1.2** and **R1.3**). Furthermore, the voting pattern shows the same distinctive phases (dotted line) as in the WL (refer to Figure 10.1 and 10.2). The *proximal* phase ( $N$  to  $N+2$ ), *medial* phase ( $N+3$  to  $N+8$ ) and the *distal* phase ( $N+9$  to  $N+13$ ). Figure 33 below shows the same data split to show the 2013/2014 cohort and the 2015/2016 cohort.

Table 16

*The Overall  $\chi^2$  Value for Each Spatial Position for 2013/2014 and 2015/2016 Cohorts*

Spatial position	Observed	Expected	$\chi^2$
$N$	58.000	81.667	6.859
$N + 1$	42.000	73.244	13.328
$N + 2$	44.000	64.288	6.402
$N + 3$	63.000	55.331	1.063
$N + 4$	59.000	47.042	3.040
$N + 5$	73.000	38.502	30.910
$N + 6$	36.000	30.413	1.027
$N + 7$	31.000	22.856	2.902
$N + 8$	23.000	16.121	2.935
$N + 9$	10.000	9.887	.001
$N + 10$	6.000	5.430	.060
$N + 11$	3.000	2.173	.315
$N + 12$	1.000	1.279	.061
$N + 13$	.000	.552	.552
<b>Total</b>	<b>449.000</b>	<b>448.783</b>	<b>69.454</b>

*Note.* The overall (i.e., positive, negative and neutral votes) observed and expected frequencies were derived from 449 participants from two different cohorts (i.e., 2013/2014 and 2015/2016).

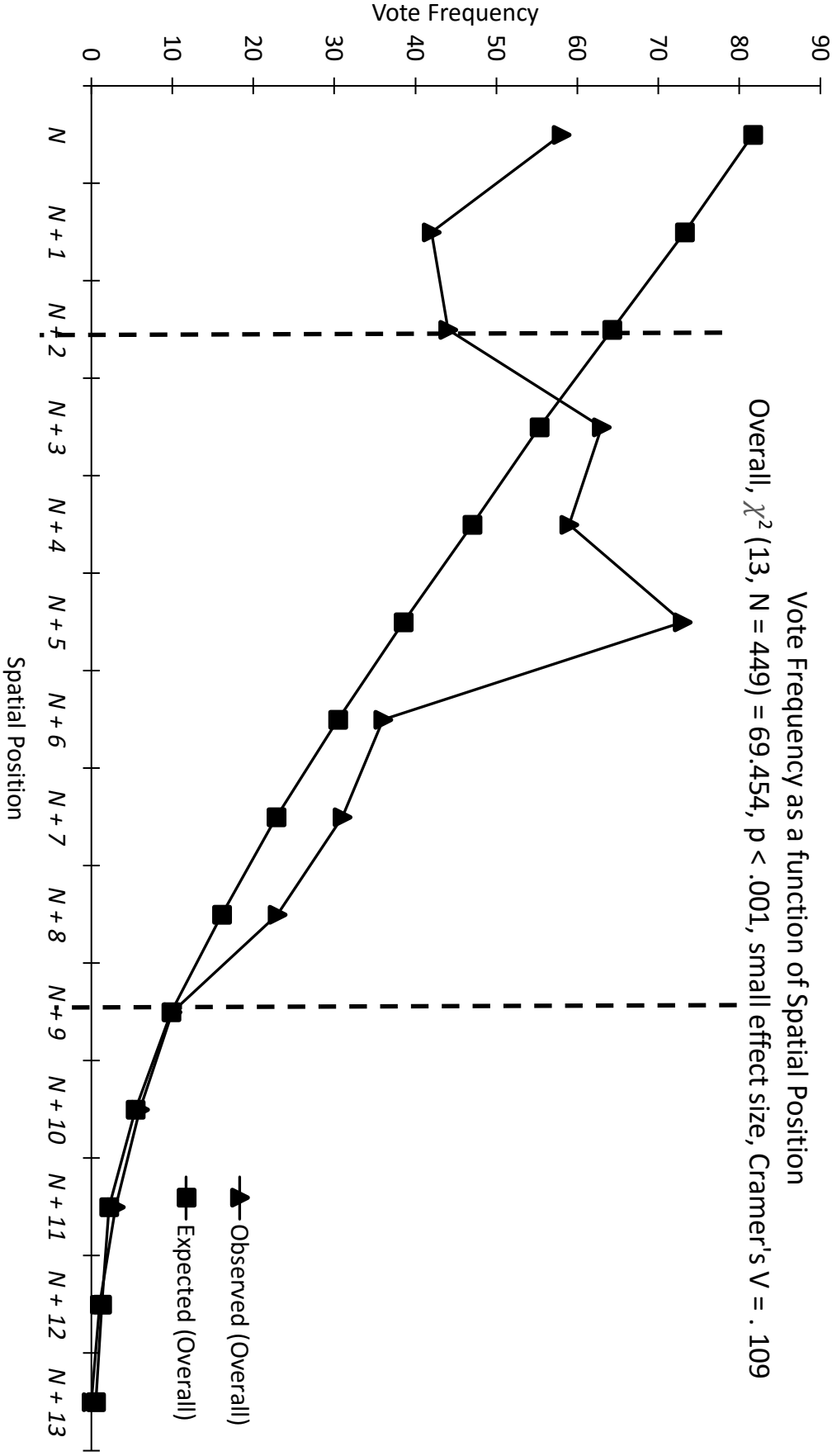


Figure 32. Overall vote frequency as a function of spatial proximity. The *proximal*, *medial* and *distal* is similar as observed in R1.2 and R1.3

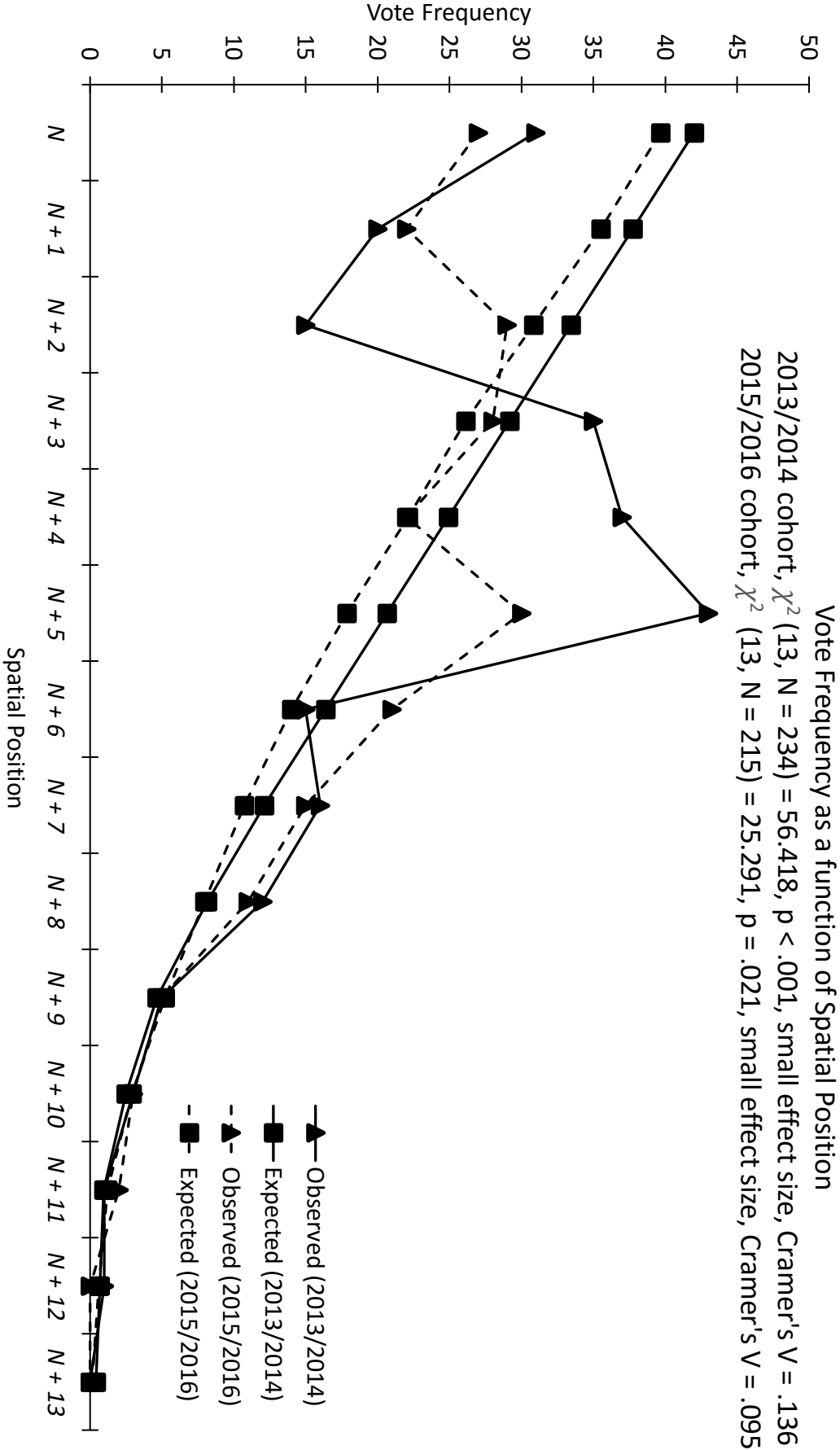


Figure 33. Overall vote frequency as a function of spatial proximity for 2013/2014 and 2015/2016 cohorts

***The effect of valence.*** In addition, the NE was further investigated to test the association with voting valence. In contrast to the WL study, this lecture study differentiated between positive and negative voting. Table 17 shows the calculated  $\chi^2$  for each valence as a function of *voter-votee* spatial position. The results show that, there was a significant association between voting and spatial position in negative voting (from 2013/2014 and 2015/2016 cohorts, see Figure 34),  $\chi^2(13, n = 211) = 50.169, p < .001$ , medium effect size, Cramer's  $V = .135$ , and it shows that there was a significant association between voting and spatial position in positive 2015/2016 cohorts, see Figure 37),  $\chi^2(13, n = 190) = 24.637, p = .026$ , small effect size, Cramer's  $V = .100$ . Whereas, for the neutral voting (i.e., as a control task, only from the 2013/2014 cohort, see Figure 42), it also showed that there is a significant association between voting and spatial position,  $\chi^2(13, n = 48) = 40.810, p < .001$ , large effect size, Cramer's  $V = .256$ .

Figure 34 shows an overall negative voting pattern (from 2013/2014 and 2015/2016), which is consistent with the voting pattern that was observed in the WL and for the overall voting pattern in this study, in which it was observed that, the participants were avoiding to cast negative votes towards other participants who were sitting the closest ( $N$ ) to them. Figure 35 below shows the same data separated for the two different cohorts. Both show significant associations between the negative vote frequency and spatial position (2013/2014 cohort,  $\chi^2(13, n = 98) = 33.839, p < .001$ , medium effect size, Cramer's  $V = .163$ . and 2015/2016 cohort,  $\chi^2(13, n = 113) = 27.476, p < .01$ , medium effect size, Cramer's  $V = .137$ ). A similar pattern was found in Figure 38 for the neutral voting, in which the participants also avoided the closest participants to receive a neutral vote. However, it shows a sudden spike for the participants who were six-spaces ( $N+5$ )

away from them. In contrast, Figure 36 below shows the overall positive voting frequencies, where the participants' votes benefitted the recipients, plotted as a function of spatial proximity of the recipient ( $N$  = neighbour). For comparison, Figure 37 shows the same data partitioned for each cohort (2013/2014 cohort,  $\chi^2(13, n = 88) = 17.494$   $p = .178$  and 2015/2016 cohort,  $\chi^2(13, N = 102) = 14.406$ ,  $p = .346$ ).

Table 17

*The  $\chi^2$  Value for Each Spatial Position Over Voting Valences for 2013/2014 and 2015/2016 Cohorts*

Spatial position	Negative	Positive	Neutral
$N$	15.438	.352	.908
$N + 1$	5.212	4.606	4.425
$N + 2$	2.286	2.950	1.321
$N + 3$	2.308	.003	.004
$N + 4$	4.054	.644	.247
$N + 5$	8.205	11.713	16.816
$N + 6$	1.586	.149	.081
$N + 7$	1.277	.139	4.179
$N + 8$	4.805	.055	.091
$N + 9$	.467	.651	.300
$N + 10$	3.389	1.296	.229
$N + 11$	.021	1.439	.138
$N + 12$	.778	.430	12.071
$N + 13$	.343	.210	.000
<b>Total</b>	<b>50.169</b>	<b>24.637</b>	<b>40.810</b>

*Note.* The  $\chi^2$  voting valences were derived from 211 participants for negative voting, 190 participants for positive voting and 48 participants for the neutral voting.

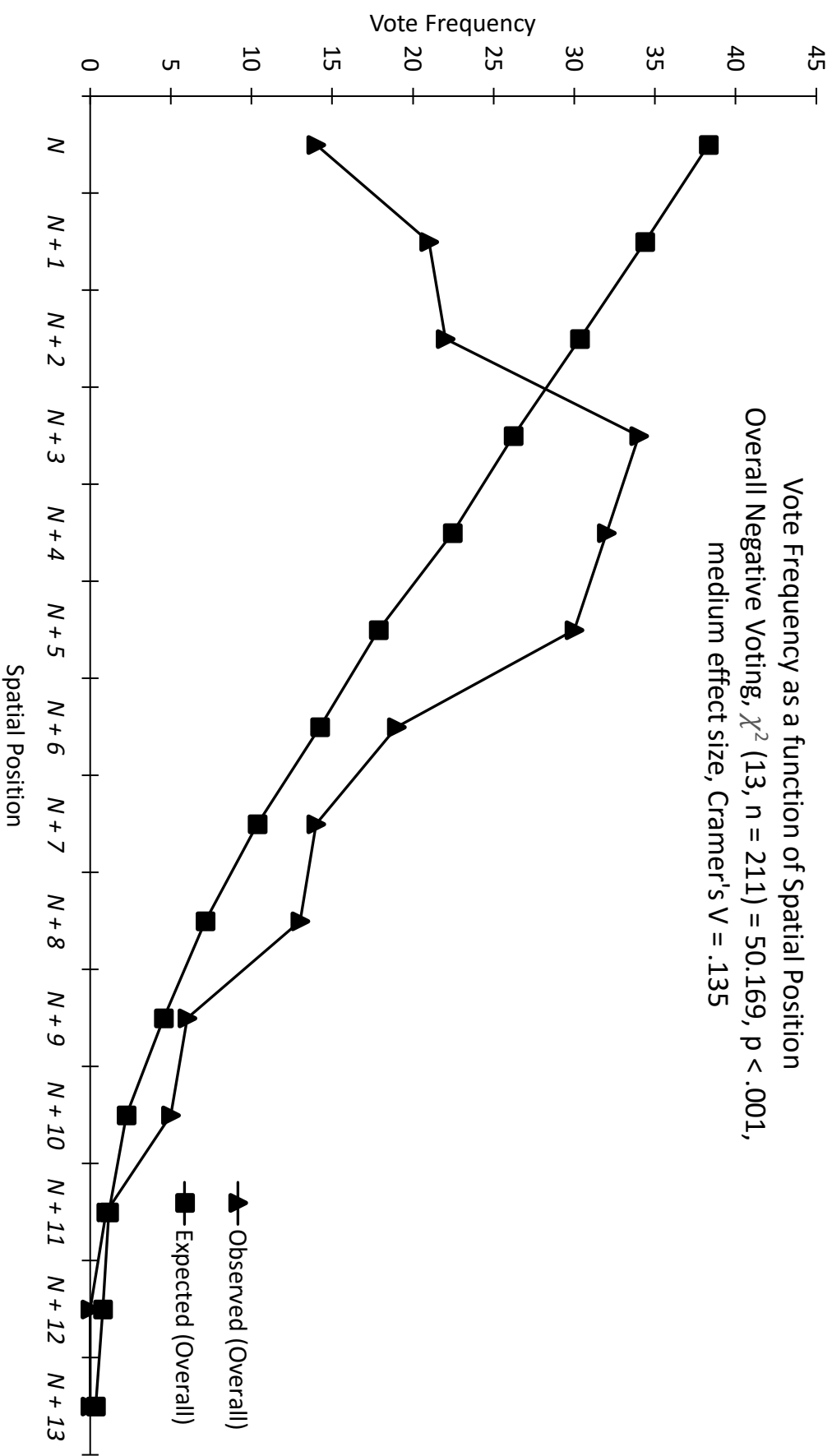


Figure 34. Overall negative voting from 2013/2014 and 2015/2016 cohorts.



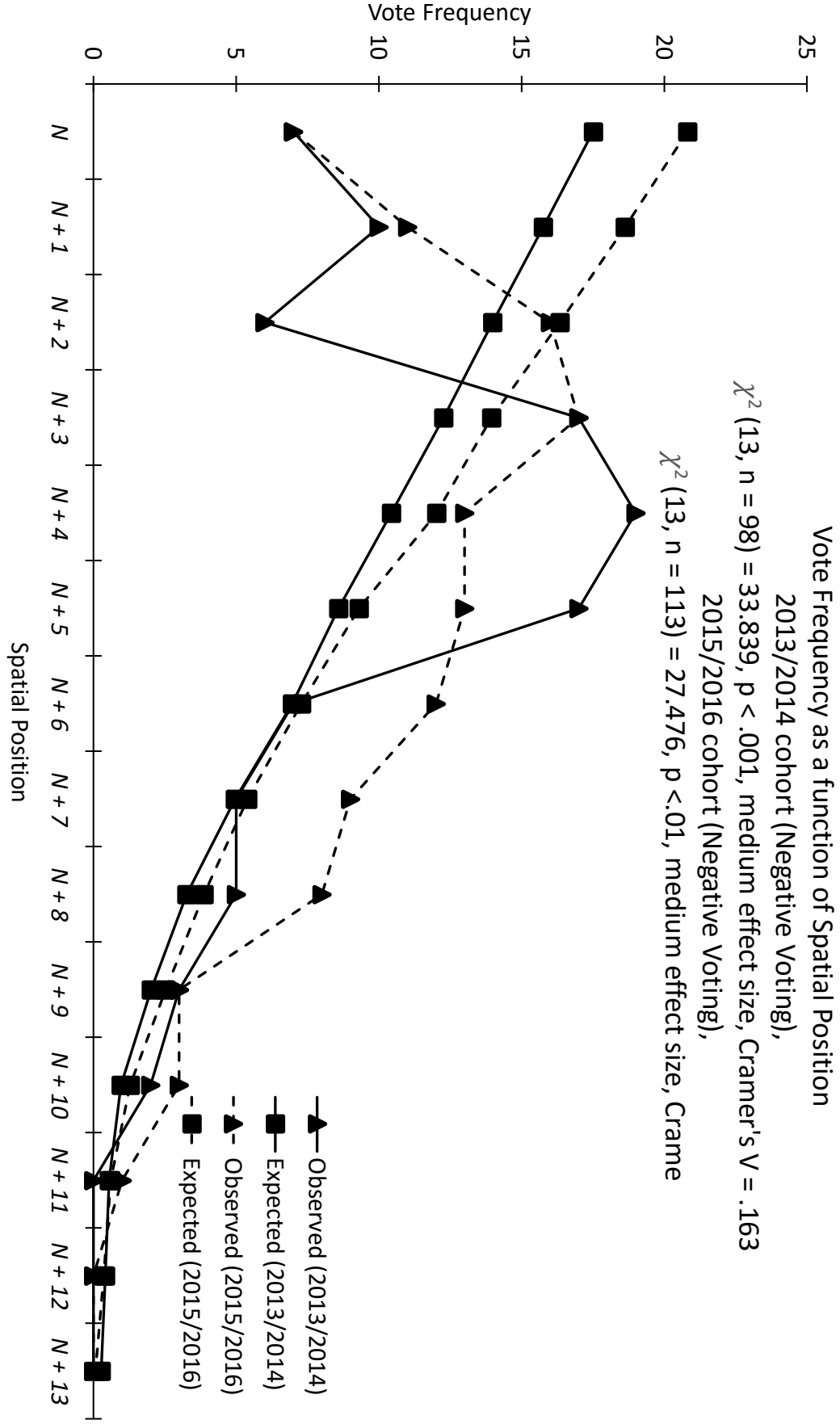


Figure 35. Overall negative voting by different cohorts (2013/2014) and (2015/2016)

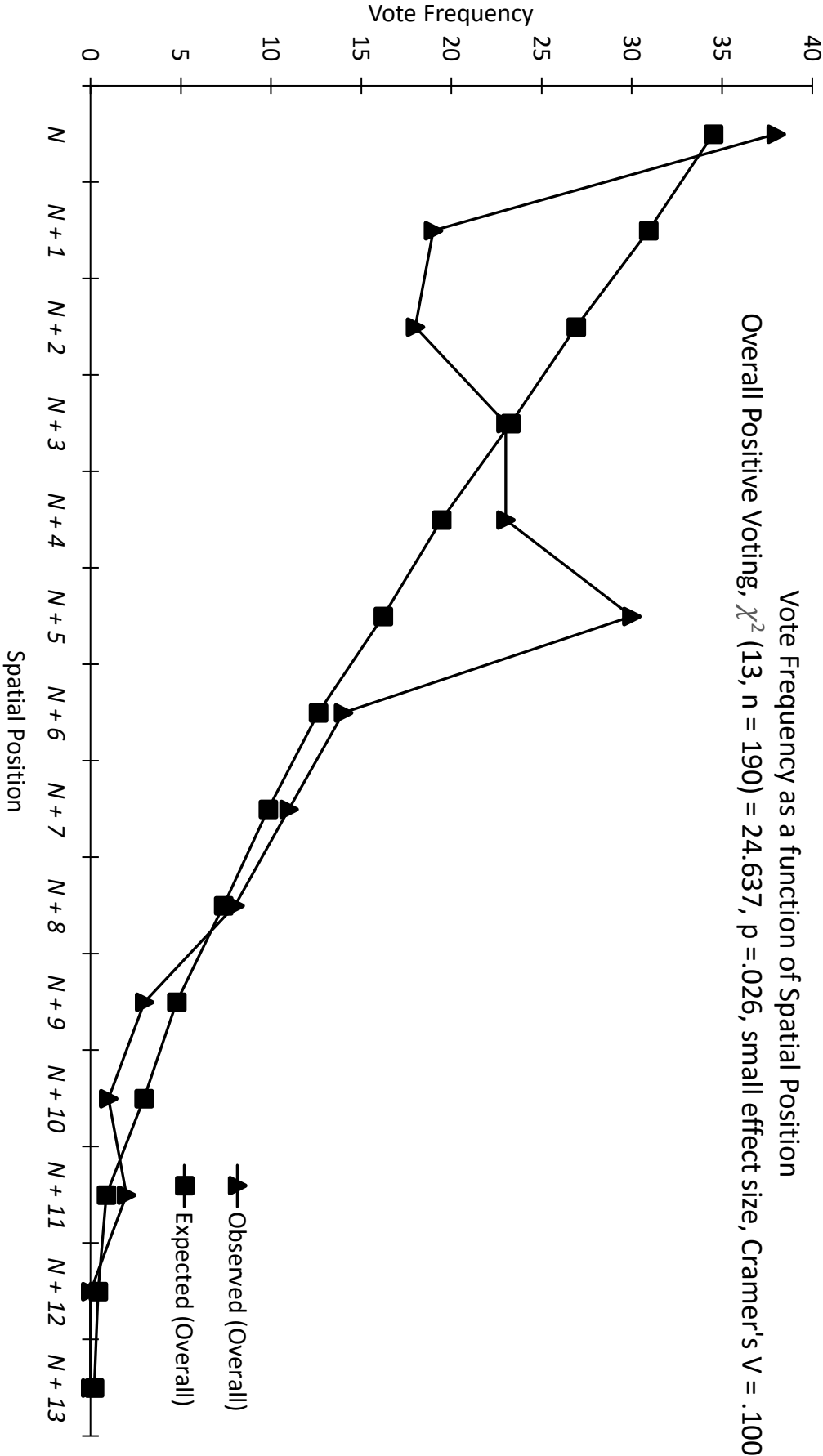


Figure 36. Overall positive voting from 2013/2014 and 2015/2016 cohorts

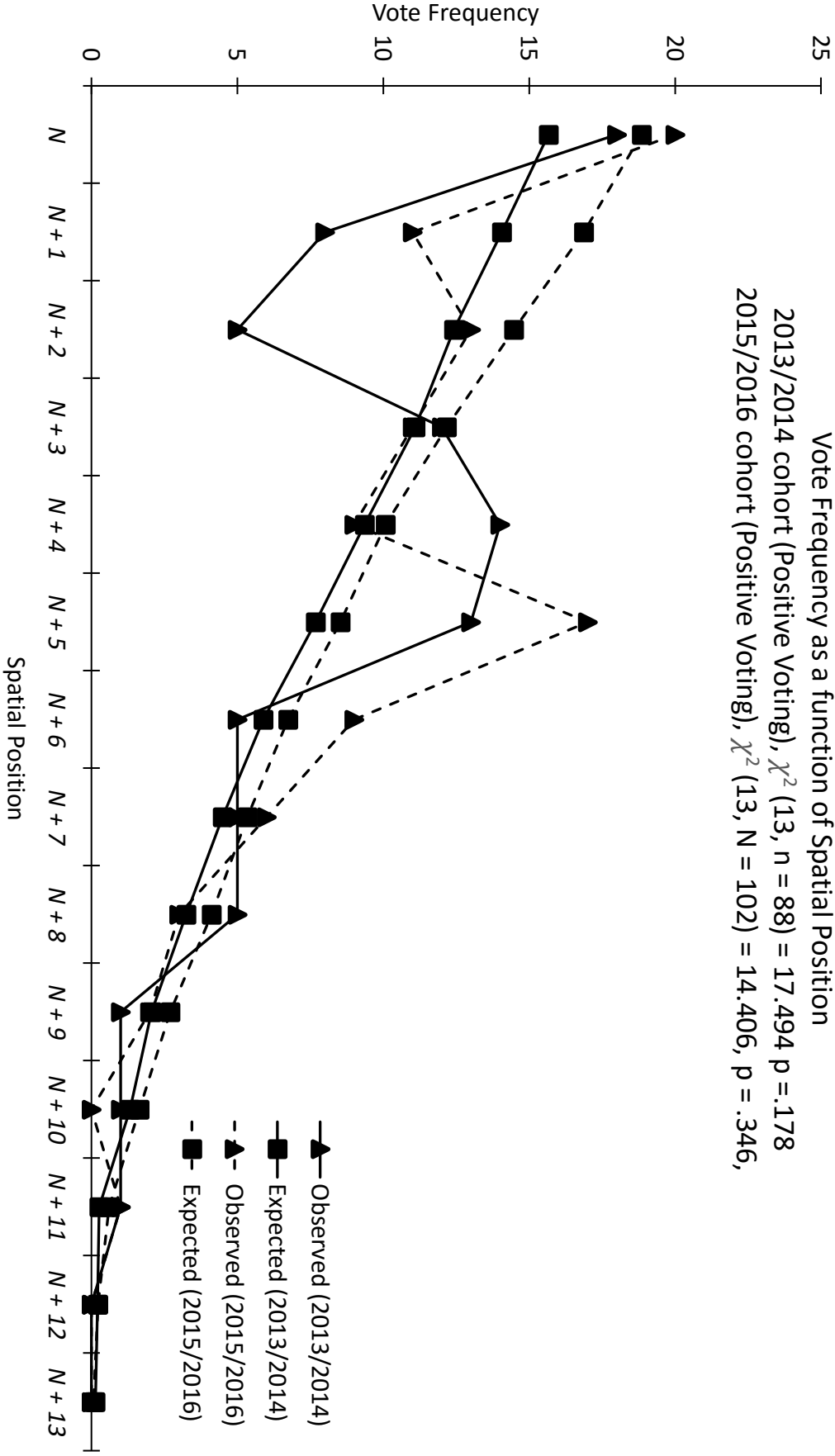


Figure 37. Overall positive voting by different cohorts (2013/2014) and (2015/2016)

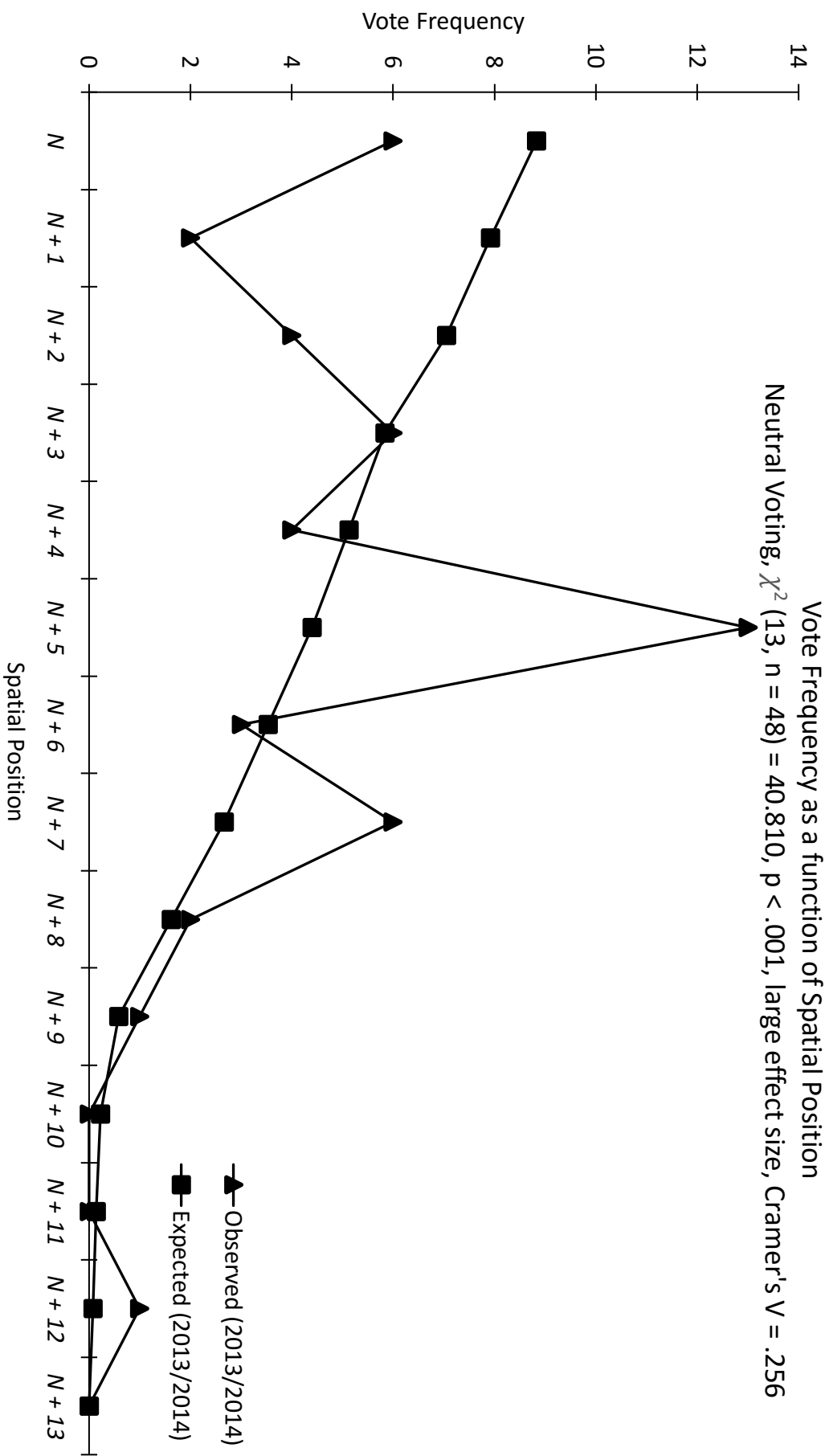


Figure 38. Neutral voting from 2013/2014 cohort

*Further observations on positive voting in the 2013/2014 lecture study.* Figure 39 and 40, below, show the positive and negative valence voting respectively for the 2013/2014 cohort only. The vote magnitude changed (either +1/+5; -1/-5), which means, at +5, the recipient received higher chances to participate in the prize draw. For votes that carried a value of +1, the  $\chi^2$  test shows that there is no significant association between the vote valence and the spatial position,  $\chi^2 (13, n = 45) = 16.575, p = .219$ . Likewise, when the vote carried a value of +5, the result shows that there is no significant association between the vote at +5 with the spatial position,  $\chi^2 (13, n = 43) = 9.748, p = .714$ . Whereas, for the negative voting in 2013/2014 cohort, Figure 40 presents the votes which carries -1 and -5. For the votes that carries -5, it lesser chances in contrast to -1, for the recipient to be nominated as the candidate for the prize draw. For the vote that contributes to -1, the  $\chi^2$  test shows that there is no significant association between the vote valence and the spatial position,  $\chi^2 (13, n = 49) = 21.972, p = .058$ . However, for the votes that gives -5, it shows that there is a significant association with the vote valence and the spatial position,  $\chi^2 (13, n = 49) = 23.962, p = .033$ , medium effect size, Cramer's  $V = .194$ .

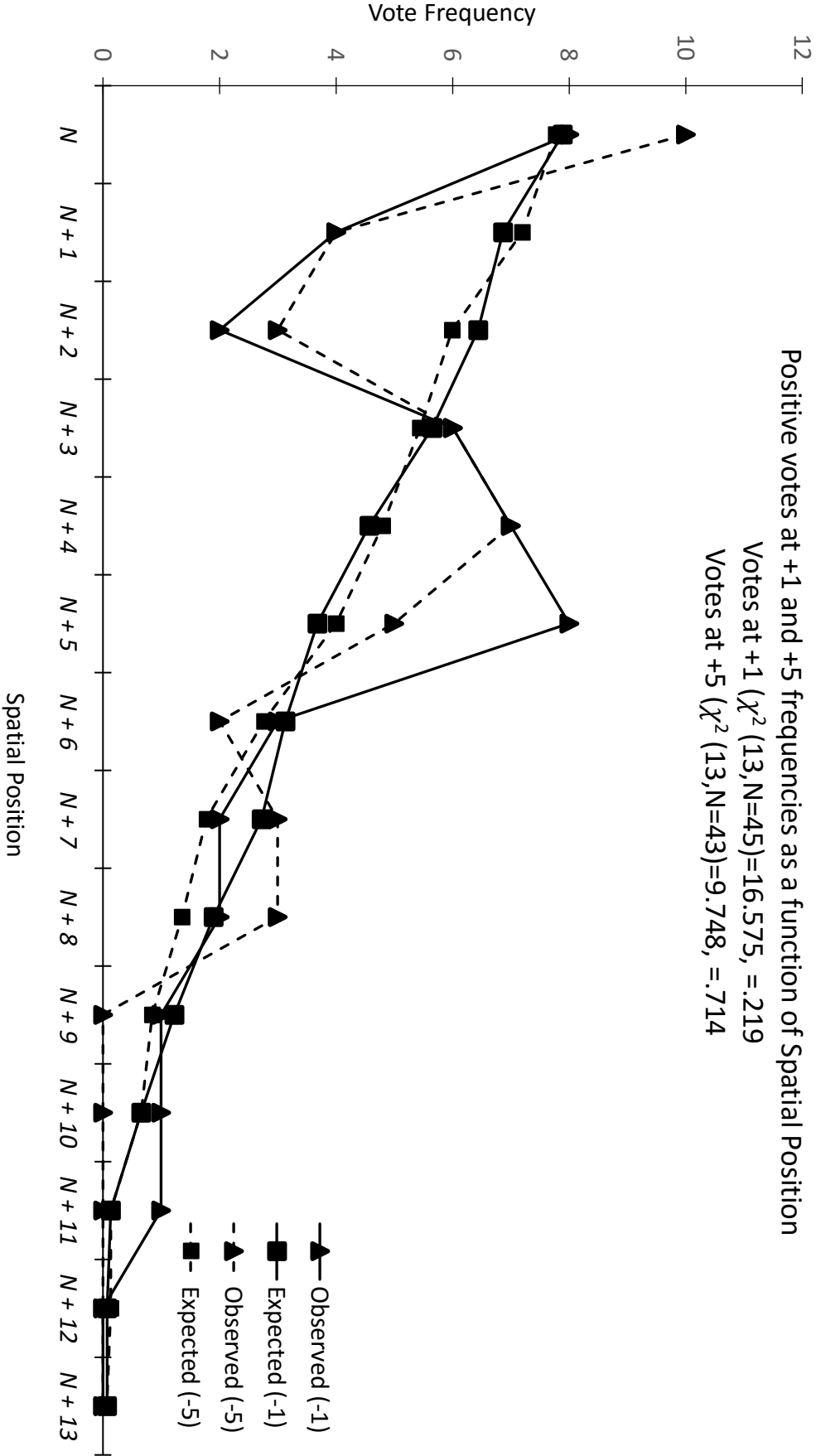


Figure 39. Positive voting at +1 and +5 valences in 2013/2014 cohort

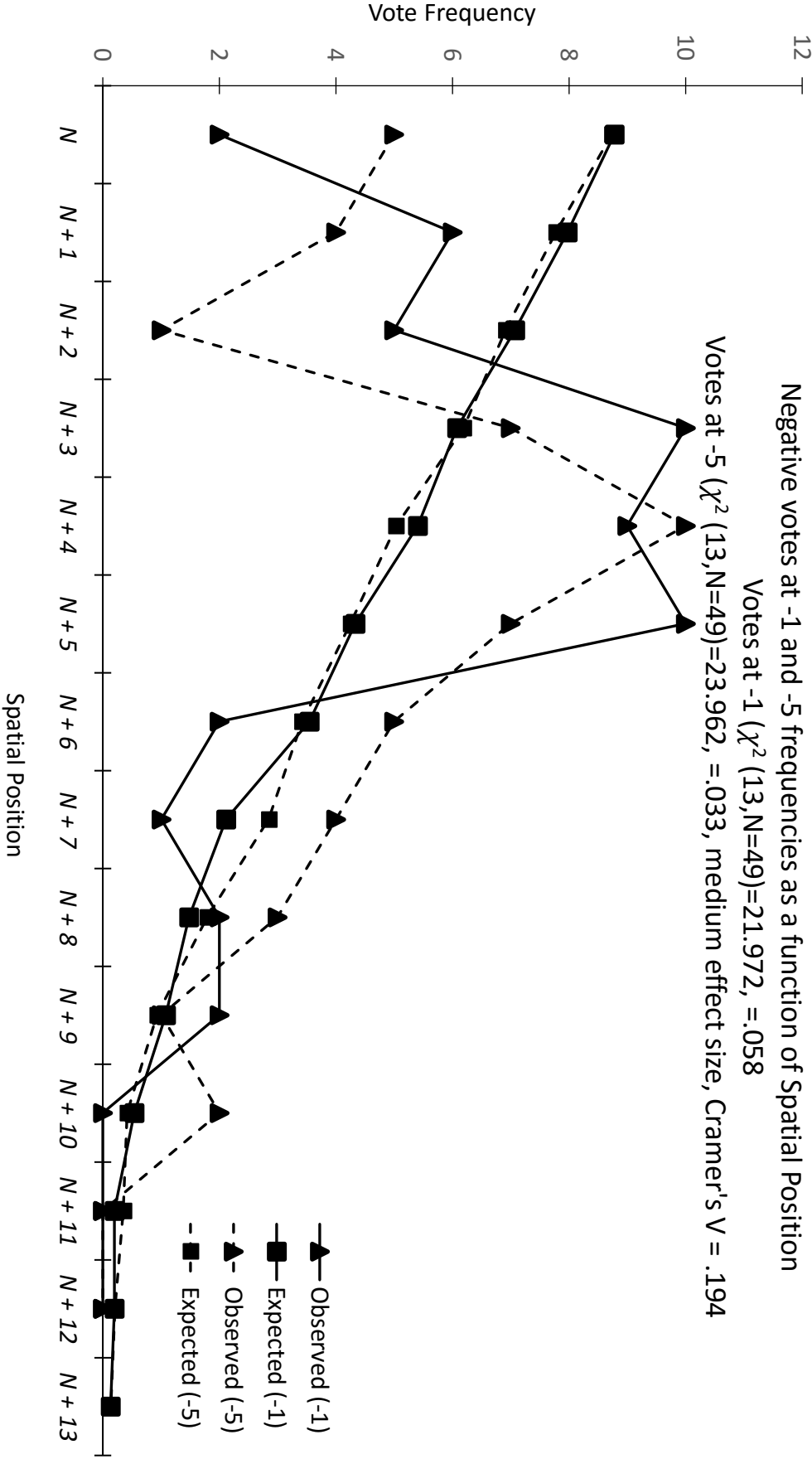


Figure 40. Negative voting at +1 and +5 valences in 2013/2014 cohort

## R2.5 Discussion

This study is important for several reasons. Firstly, it replicates the NE that was originally shown in the voting patterns of contestants on the WL TV game-show (Goddard et al., 2011; Goddard, 2012; Goddard et. al., 2013). The contestants on the WL game-show significantly avoided doing a nasty thing to their direct spatial neighbour(s), by not voting for them as the weakest player. However, the replication in the lecture study was in some ways even more surprising because, whereas the votes cast in the WL were open and public, in the lecture study votes were closed and private.

A limitation of the WL field experiment was the lack of control over key variables and conditions. This study, however, included the scenario where the students' selection of a peer carried a positive outcome rather than just the negative outcome in WL. When the student participants were asked to do a nice thing towards someone along the same row, they tended to favour their direct spatial neighbour(s) when it was seen as conferring a positive benefit. This is a particularly important finding as it shows that the *neighbour effect* (NE) is not simply a bias in making spatial selections per se, rather it changes with the positive/negative valence of the outcome.

At a cognitive level the voting behaviour in the WL and this study can be considered in terms of them utilising two sources of information to make their decision and cast their vote. Firstly, there is the primary source, of open public information that everybody has access to and can be used to make a rational informed decision. For example, in the WL this would be the performance of the contestants answering questions.



If one contestant repeatedly makes errors responding to questions but all their fellow contestants make no errors, then the decision is simple. However, when this information is equivocal, uncertain and ambiguous, then the secondary source of information based on intuition is used. This is private, individual, subjective and prone to bias. This secondary source of information comes to the fore, when the primary source is not reliable. It suggests that in the WL, and in this lecture study, that the NE comes about when the contestants and student participants have to make decisions when there is a limited primary source of information on which to base their judgement. Under these conditions, error prone, biased intuition dominated objective decision criteria. This NE is an underlying bias in decision-making that is part of this secondary source of information. This is akin to Kahneman (2011) fast and slow thinking in decision-making and it concurs with bounded rationality theory (Simon, 1955), where the decision maker's rationality is limited by information processing capacities and access to information.

When the valence of the vote is negative it can be considered as in some way punishing the receiver. This places the voter in a dilemma especially if they operate a 'do-no-harm' principle (Baron, 1995), where participants will be reluctant to harm others. It seems that they avoid meting out punishment to their direct-neighbour(s) possibly through fear of retaliation. On the other hand, when the valence of the vote is positive it can be considered as a kind of gift to be bestowed on a valued individual (e.g., Bell, 1991). In this case the NE disappears and the neighbour is once again favoured. Nevertheless, the effect was asymmetric being far stronger in the negative rather than the positive valence scenario. This suggests that the participants' fairness norm was stronger in the negative valence condition compared to the positive valence condition, whereby the participants were

favouring the negative recipient(s) by avoiding doing bad thing towards them (Leliveld, van Beest, van Dijk, & Tenbrunsel, 2009).

As opposed to the WL TV game-show, that involved with public open voting and high stakes, in this study, the NE was also tested with comparatively low stakes. As the voting activity incorporated positive as well as negative voting valences, it concurs the observation by Kocher et al. (2008) and Fehr et al. (2014) which suggested that the stakes level does not affect the decision-makers behaviour. In which, for this study, the behaviour of avoidance to penalize their direct-neighbour(s) in a negative voting valence regardless of the stakes size were consistent.

Thus, it suggests that the NE occurred as an implicit bias in decision-making, probably working at an unconscious level and possibly arising as a reluctance to engage in actions that could potentially result in conflict with those most nearest to them. Therefore, both predictions relating to proximity bias and valence were supported.

However, studies have consistently shown that, the NE is a bias in decision-making and question has been raised, if this bias existed as result of conflict avoidance to those who nearest in proximity (at  $N$ ). Hence, the next study seeks to address the robustness of the NE, if the participants were given the choices to either cooperate or defect those who is closest in proximity (at  $N$  and  $N+1$ ) in a form of experimental game.

## Chapter R3

### Testing the Neighbour Effect using Prisoner's Dilemma

#### R3.1 Chapter Outline

R2 demonstrated another compelling *neighbour effect* (NE) that emerged in a field experiment set in a lecture theatre. Its success prompted the design of this experiment. If the NE can be demonstrated in a simple voting behaviour on WL and in a lecture could it also have an impact on a well-established classic scenario like the Prisoner's Dilemma (PD).

This chapter is presented in four sections: a general introduction that focuses on the PD perspective to understand the NE, the elaboration of the methods, results and discussion.

R3.1 Chapter outline

R3.2 Introduction

R3.3 Methods

R3.4 Results

R3.3 Discussion

**Abstract.** The *neighbour effect* (NE) is a bias in social decision-making exhibited by actors that (usually) favours their direct-spatial-neighbours rather than non-neighbours. The NE was tested using participants seated in a lecture theatre ( $N = 229$ ) by getting them to play a closed form of the Prisoner's Dilemma game. Participants had to choose whether to cooperate or defect with either their direct-spatial-neighbours or non-neighbours. The result showed that, the participants were significantly more cooperative with their lateral (left/right) direct-spatial-neighbour(s) than non-neighbours. However, the participants became significantly less cooperative with their longitudinal (front/back) direct-spatial-neighbour(s) than their non-neighbours,  $\chi^2(1) = 7.376$ ,  $p = .007$ . Therefore, NE is not just a bias based on simple spatial proximity but is based on the relative lateral/longitudinal positioning of actors.

### R3.2 Introduction

This chapter discusses the exploration of the *neighbour effect* (NE) in a form of PD game format to understand the extent of the NE in a cooperative social interaction. The participants (i.e., players) were prompted to choose either to cooperate or defect with another player with the goal to maximize their utility: to earn as many points as possible.

Demonstration and the replication studies in **R1.2 and R1.3** established that the NE is a profound bias in decision-making. The contestants of the WL TV game-show significantly avoided penalizing their direct-neighbour(s) when they were called to identify one of the contestants to be the WL (Goddard et al. 2011; Goddard, 2012). In addition, **R2** demonstrated the effect of vote valence (i.e., negative and positive voting). A vote that carried negative consequences to the recipients (i.e., another recipient) showed the consistent proximity-voting pattern as observed in the observations of WL in **R1.2**. However, a positive vote for a recipient demonstrated that the *voter* favoured the closest recipients. Hence, this study was designed to contest the notion that the NE is a bias in decision-making when the obvious proximity exists.

Gardin, Kaplan, Firestone and Cowan (1973) suggest that, there was a significant association between positive cooperation with proximity when the participants were seated at proximal side-by-side when the visibility was blocked. Their study used PD to examine the effect of the seating arrangement and other attributes such as, visual contact, attitude and approach-avoidance tendencies. In their method, 80 male participants were assigned randomly into pairs up to 10 pairs for each condition (i.e., four conditions of the payoff

values: 5/5, 0/6, 0/0, 1/1). The game was repeated in 50 trials for each condition. In the session, the participants were assigned to be seated either side-by-side or across-table seating condition. During the session, dividers were used as barriers in some of the conditions. Prior to begin the session, the participants were instructed to accumulate as many point as possible and there was no constraint to confront each other either in extreme cooperation or competition. At the end of the session, the participants were asked to answer a questionnaire on their response to the session, and the earning was distributed (10 points = \$0.10) and participants were debriefed. Their study suggested that, the participants were more likely to cooperate with the absence of the barrier when the participants were seated across, but less likely to cooperate when they were seated next to each other. However, a reverse pattern was observed when there was a barrier. The participants who were seated next to each other tended to cooperate more than those who were seated across from each other. In the other condition, as in approach avoidance and interpersonal attitude, in both conditions where the participants were seated across from each other, with no barrier, showed a positive pattern. Gardin et.al's study was important as it shows an effect of seating arrangement on cooperative behaviour. However, they also noted that the participants' responses to questionnaires neglected any mention of the seating arrangement. This suggests that the effect of the participants seating was endogenous and implicit.

De Heus, Hoogervorst and Van Dijk (2010) investigated the effect of positive and negative valence payoffs in Prisoner's dilemma and the Chicken game. They argued that, to understand risky choices in social dilemmas, either to cooperate or defect, it was vital to have clear 'risky' options. Participants (N=198) were recruited for a 2 x 2 game format

(see Figure 41 that shows the pay-offs structure presented to the participants) where they were presented with either the Prisoner's dilemma or the Chicken game with framing between gain versus loss (valence). The participants were allocated to separate cubicles equipped with PCs, they were instructed that all communication, to the experimenter and other participants, must be through the PC, hence proximity was not measured in relation to understanding the valences in social dilemmas. Participants completed questionnaires to measure risk orientation and were informed that they would be involved in a game with another participant to win lottery tickets, which they would give them a chance to win a voucher worth €20 (approximately £17). The participants played a one-shot game (i.e., Prisoner's dilemma or Chicken game) where they chose between A (cooperation) or B (defection).

Differences were found between how the two dilemmas were played out and this was attributed to the clearer payoff structure in the chicken game with regard to risk. Participants with a tendency for risk showed greater willingness to take risks by playing B (defection) in the Chicken game relative to Prisoner's dilemma.

The effect of valence was clearer in the Chicken game than in the PD, where the participants perceived a defection was riskier choices as in the Chicken game. In the Chicken game, the positive valence (gain) was considered as a cooperative choice than in negative valence (loss) compared to PD game where it showed insignificant framing effect towards both valences. Hence, from this study, it was suggested that, to understand risk taking behaviour, clearer options and alternative valences of choices should be presented.

		Prisoner's dilemma		Chicken game	
		Other		Other	
		C	D	C	D
Gain frame	Self C	2, 2	0, 3	2, 2	1, 3
	Self D	3, 0	1, 1	3, 1	0, 0
		Other		Other	
		C	D	C	D
Loss frame	Self C	-1, -1	0, -3	-1, -1	-2, 0
	Self D	0, -3	-2, -2	0, -2	-3, -3

Figure 41. The gain-frame and loss-frame pay-off structures for the Prisoner's dilemma and the Chicken game. The pay-off matrix is reprinted from "Framing prisoners and chickens: Valence effects in the prisoner's dilemma and the chicken game" P. de Heus, N. Hoogervorst and E. van Dijk, 2010, *Journal of Experimental Social Psychology*, 46, 737-742. Copyright (2010) by Elsevier Inc.

In a more recent study, McCall and Singer (2015) suggested that, there is a significant relationship between proxemic images and unfairness behaviour. In their study, they investigated four different tasks (i.e., Prisoner's Dilemma game, memory task in virtual environment, punishment task and post-task questionnaire). In the PD game, the participants were asked to play the game on a computer and each player was represented as an avatar. The task was played by three players (i.e., the participant become an avatar, two other players were avatars created by the experimenter) in each session where prior to the session, the first player was given 10 monetary units, in which the player can decide either



to keep or to transfer to the second player. Participants were led to believe that one of the players (e.g., the player B) was always the first player and for each round, the other players were chosen to be the second player. The next task was also tested in a virtual environment for a memory task where the participants were told that, towards the end of the task, they will be quizzed to test their memory. For this task, the participants were geared up with the head-mounted display and they were lead into the virtual world. In the virtual environment, two other players were stood in the room and the participants were led to believe that the other players were located in separate rooms.

The task involved looking at images and green dots for approximately three minutes. In the third task, they were asked to play monetary punishment whereby, for each contribution made, the other player will be punished by losing their monetary unit. The final task, was to rate their liking towards of the other two players. To ensure that there was no manipulation, McCall and Singer validated that, the fairness manipulation was perceived as a salient behaviour where they run a simple t-test to confirm the fairness. Another prominent finding in their study was the pattern of gaze: the high punisher tended to lock gaze with the other two players more than the low punisher. Yet, it could be argued, that the pattern observed in this study existed merely due to lack of real interaction between participants and the players due to the context of the study done in a virtual world. Hence, they highlighted that, for future research to explore whether the participants were able to distinguish differences between the known proxemic pattern, which could affect the social interaction and to further the investigation with nonverbal cues.

Hence, this study was designed to challenge the limitation of the aforementioned studies and also to understand the proximal effect in NE. It is believed that, such behaviour could also exist in a perception of self-interest, the ideal concept could be derived from the Prisoner's Dilemma (PD). In PD, the concept of rationality could affect an individual's choice when presented with options to either cooperate or defect with each other.

**Aim, Objectives and Hypotheses.** The aim for this study was to challenge the notion of the NE as a spatial bias in the context of the PD's game. It is assumed that if players are rational then they should seek to maximize utility and play the Nash equilibrium by defecting, the unique solution for the game. However, if, neighbours are swayed by the NE then they might opt to cooperate. Hence, to achieve the aim of this study, the objectives of this study were;

1. To test whether the endogenous factor of the NE (derived from previous studies) extends to choices in PD.
2. To test the association between cooperation and defection (i.e., endogenous factors) in NE.

Therefore, several assumptions were made based on the PD model (Figure 42) to understand the *voter* dilemma, which in this context is the player *Y*, while the recipient is player *X*. The payoff matrix was extended from the original version (see Figure 2) by Tucker (1983) and the pay-off matrix in De Heus et al., 2010 (see Figure 41), in which it follows the rank of  $T > R > P > S$ .

		Player <i>X</i>	
		heads (H)	tails (T)
Player <i>Y</i>	heads (H)	(3, 3) <b><i>RR</i></b>	(0, 5) <b><i>ST</i></b>
	tails (T)	(5, 0) <b><i>TS</i></b>	(1, 1) <b><i>PP</i></b>

Figure 42. The pay-off matrix for the coin's game. The first number in the brackets gives player *Y*'s payoff and the second number gives player *X*'s.

**Assumption 1. (Reward, *R*)**

If player *Y* decides to play H with player *X*, with assumption that the player *X* plays H, the accumulate point is six.

**Assumption 2. (Temptation, *T*)**

If player *Y* decides to play T with player *X*, with assumption that the player *X* plays H, the accumulate point is five.

**Assumption 3. (Punishment, *P*)**

If player *Y* decides to play T with player *X*, with assumption that the player *X* plays T the accumulate point is two.

**Assumption 4. (Sucker, *S*)**

If player *Y* decides to play H with player *X*, with assumption that the player *X* plays T, the accumulate point is five.

Therefore, in this pay-off matrix, the optimal strategy for the player *Y* is to play T in which, for any situation given with assumption that, the player *X* might defect by playing T, the player *Y* will always be better off and earn better points, five (SW) and one (SE) respectively. In addition, if the player *Y* trusts the player *X*, the dominant strategy is by playing H. Thus, given four assumptions of the strategies that can be deployed by the player *Y* with regards the proximity and outcome from the payoff, the hypotheses were;

1. The null hypothesis ( $H_0$ ) suggested that, there is no association between the player *Y*'s strategy and spatial proximity.
2. The first alternative hypothesis ( $H_1$ ) predicted that, the player *Y* will be more likely to play H with player *X* who is closer in proximity.
3. Second hypothesis ( $H_2$ ) suggested that, the player *Y* will be more likely to play H with player *X* who is closer in proximity with high visibility.

Thus, the following section presents the method in this study to test the association between the NE and the spatial proximity in the context of PD game.

### R3.3 Method

**Participants.** First-year Psychology undergraduates (N = 229; Female = 182, Male = 30, Undisclosed = 17; 2014/2015 cohort; age between 18 to 46 years old), were recruited at their orientation lecture during their first induction week at the University (refer to **M1.4** p. 67, for the setting and participants). Table 18 presents the participants' gender and age group recruited for this study.

Table 18

*The Participants' Demographics (n=229) for Cohort 2014/2015*

Characteristic	n	%
Gender		
Female	182	79.476%
Male	30	13.100%
Undisclosed	17	7.424%
Age		
18 - 24	214	93.450%
25 - 34	4	1.747%
35 - 44	5	2.183%
45 - 54	2	0.873%
Undisclosed	4	1.747%

**Note.** 67% of the participants were 18 years old. Participants were informed that they have a right not to disclose their gender and age.

**Procedure.** For this study, the participants were asked to participate in an activity, which required them to play a classic *Coins game* (a simple version of PD). The task took approximately 10 minutes to complete including introduction and debrief. The rules of the games and the procedure were explained in general in **M1.4**.

The goal of the *Coins Game* (CG) activity required participants to score as many points as possible in three simple games. The participants had to choose either H or T (denoting Heads or Tails on a coin) by circling the answer on their response sheet for *Game 1* and *Game 2*. For *Game 3* in which the participants were required to write either H or T to play against another player *X*. They were shown the four permutations for cooperation/defection (refer to Figure 42 for the CG payoff matrix) and they had to decide the best strategy to play (participants were required to score as many points as possible when playing against another player for the three games, and the scores were averaged out). The task was performed secretly without the knowledge of the other participant (player *X*). Table 19, shows the original payoff matrix (refer to Appendix C1 to C6) presented to the participants.

Table 19

*Original Table Score Presented in the Response Sheet*

<b>you</b>	<b>them</b>	<b>your score</b>	<b>their score</b>	<b>Total</b>
<b>H</b>	H	<b>3</b>	3	6
<b>H</b>	T	<b>0</b>	5	5
<b>T</b>	H	<b>5</b>	0	5
<b>T</b>	T	<b>1</b>	1	2

**Note.** This is the original table that was given to the participants on the response sheet (refer to Appendix C)

For the task, the participants (always player *Y*) played three games with another participant (player *X*) potentially seated at, (a) the same block of the lecture theatre and in the same row of seats, (b) the same block of the lecture theatre and one row at the front and back and (c) within the lecture theatre, either as their direct-neighbour(s) or their two-spaces away neighbour(s).

Six different versions (refer to Appendix C1 to C6) were randomly allocated (in sealed envelopes) to the participants. Table 20 shows the distribution of the completed form (response sheet) when the participants completed all the required questions (two screening questions, *Game 1*, 2 and 3, refer to Appendix C1 to C6). The collected data were counted and coded accordingly based on the type of the question (refer to **M1.4**)

Table 20

*The Distributions of the Coin Game Response Sheet*

Version (v)	Participants
1	42
1r	40
2	39
2r	37
3	34
3r	37
<b>Total</b>	<b>229</b>

**Note.** Each version indicates the player *X* position for *Game 1* and *Game 2* (refer to Appendix C1 to C6 for the full versions).

The following diagrams (Figure 43, 44 and 45) illustrate the three versions of the spatial arrangements (lateral, longitudinal, and their combination) and their counterbalanced mirrors.

**Version 1**

**GAME 1:-** You are playing the ‘coin game’ with the people one space away from you, to your left and right (your direct-neighbour(s)).

**How will you play? CIRCLE ONE** **H** **T**

**GAME 2:-** You are playing the ‘coin’ game with the people two spaces away to your left and right (your next-door-but-one neighbours)

**How will you play? CIRCLE ONE** **H** **T**

**Version 1r**

**GAME 1:-** You are playing the ‘coin’ game with the people two spaces away to your left and right (your next-door-but-one neighbours)

**How will you play? CIRCLE ONE** **H** **T**

**GAME 2:-** You are playing the ‘coin game’ with the people one space away from you, to your left and right (your direct-neighbour(s)).

**How will you play? CIRCLE ONE** **H** **T**

184	185	186	187	188
146	147	148	149	150
109	110	111	112	113
72	73	74	75	76
35	36	37	38	39

*Figure 43.* Diagram of version 1 and version 1r with the seat arrangement in the lecture theatre. Consider Player *Y* in seat number 111, receiving version 1. For *Game 1*, Player *Y* has an option to play against Player *X* seated either at 110 or 112. In *Game 2*, Player *Y* can play against Player *X* seated either at 109 or 113. Version 1r is the counterbalanced mirror.



**Version 2**

**GAME 1:-** You are playing the ‘coin game’ with the people one space away from you, to your left and right, front and back (your direct-neighbour(s)).

**How will you play? CIRCLE ONE** **H** **T**

**GAME 2:-** You are playing the ‘coin game’ with the people two spaces away to your left and right, front and back (your next-door-but-one neighbours)

**How will you play? CIRCLE ONE** **H** **T**

**Version 2r**

**GAME 1:-** You are playing the ‘coin game’ with the people two spaces away to your left and right, front and back (your next-door-but-one neighbours)

**How will you play? CIRCLE ONE** **H** **T**

**GAME 2:-** You are playing the ‘coin game’ with the people one space away from you, to your left and right, front and back (your direct-neighbour(s)).

**How will you play? CIRCLE ONE** **H** **T**

184	185	186	187	188
146	147	148	149	150
109	110	111	112	113
72	73	74	75	76
35	36	37	38	39

*Figure 44.* As Figure 43 above. In version 2, Player *Y* has an option to play against Player *X* seated either at 110, 112, 148 or 74 in *Game 1*. In *Game 2*, Player *Y* as an option to play against Player *X* seated either at 109, 113, 186 or 37. Version 2r is the counterbalance.

Figure 45 as Figure 43 illustrates the possible option opted by participants (Player Y) when they received the instruction for version 3 and 3r.

<b>Version 3</b>		
<b>GAME 1:-</b> You are playing the ‘coin game’ with the people one space away from you, to your front and back (your direct-neighbour(s)). <b>How will you play? CIRCLE ONE</b> <div style="float: right; text-align: center; width: 150px;"> <b>H</b>                      <b>T</b> </div>		
<b>GAME 2:-</b> You are playing the ‘coin game’ with the people two spaces away to your front and back (your next-door-but-one neighbours) <b>How will you play? CIRCLE ONE</b> <div style="float: right; text-align: center; width: 150px;"> <b>H</b>                      <b>T</b> </div>		

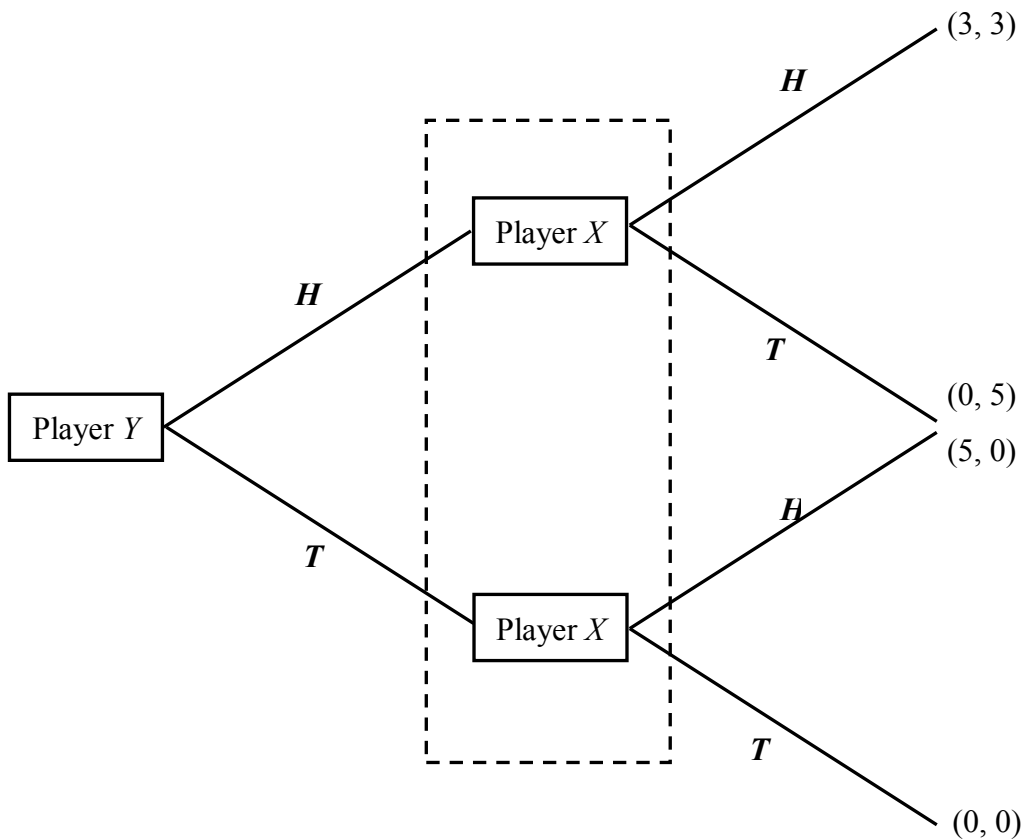
  

<b>Version 3r</b>		
<b>GAME 1:-</b> You are playing the ‘coin game’ with the people two spaces away to your front and back (your next-door-but-one neighbours) <b>How will you play? CIRCLE ONE</b> <div style="float: right; text-align: center; width: 150px;"> <b>H</b>                      <b>T</b> </div>		
<b>GAME 2:-</b> You are playing the ‘coin game’ with the people one space away from you, to your front and back (your direct-neighbour(s)). <b>How will you play? CIRCLE ONE</b> <div style="float: right; text-align: center; width: 150px;"> <b>H</b>                      <b>T</b> </div>		

184	185	186	187	188
146	147	148	149	150
109	110	111	112	113
72	73	74	75	76
35	36	37	38	39

*Figure 45.* As Figure 43 above. In version 3, Player Y has an option to play against Player X seated either at 148 or 74 in *Game 1*. In *Game 2*, Player Y can play against Player X seated either at 186 or 37. Version 3r is the counterbalance.

**Expected Frequencies for playing *H* and/or *T*.** Given the assumption that the probability of the player *Y* plays *H* with assumption that, they strategized for the optimal strategy to collect six points, there is a probability that player *X* will defect by playing *T*. Figure 46 illustrates the pathway of player *Y* strategies to achieve the optimal payoff.



*Figure 46.* Illustrates player *Y* pathway to play against player *X*. To optimize own gain, player *Y* has one optimal strategy, provided player *Y* thinks that player *X* will cooperate and vice versa.

### R3.4 Results

The results are in two parts. First, the analysis to test the NE is presented and then the second part tested the association between NE and the direction of spatial relationship. Given the unequal distribution of gender described in the lecture based voting activities previously (see p 168) gender was not studied as an independent variable in this section either (F=182; M=30 of those that expressed gender). In the screening questions, participants were asked to demonstrate their understanding of the game's rules, which showed that the majority (from fairly well and well) of the participants (79.039%) out from 229 participants expressed that, they understood the instruction either fairly well or well in comparison to 2.183% responded that, they had either not understood the game and instruction at all or not very well. The participants' knowledge on the activity by response rate were presented in Figure 47. In general, the majority of the participants were well-informed on how to play the CG. In addition, the participants were asked on their choices if, (a) the opponent played H and (b) the opponent played T. The questions were asked to understand their judgement on how they would respond to maximize the overall points to score. Figure 48 shows how the participants responded. The rational strategy was that they should play T in both situations to maximize their payoffs (see Table 19, for the payoff table). By going through these questions it was shown to the participants explicitly what the best strategy was to play the game. These questions tested that they had understood this explanation. The findings showed that the participants (Player *Y*), will choose to play T regardless the strategy played by Player *X* (see Figure 48).

Based on the choices made by each individual, it was assumed that maximizing the payoffs would be the rational decision in a situation either to defect or cooperate with another individual whilst performing the choices privately (Gottman, 2011). Further analysis shows that, 89.083% of the participants were consistent in their choices in both *Game 1* and *Game 2*, where 93.627% of the participants chosen to play T in both scenarios.

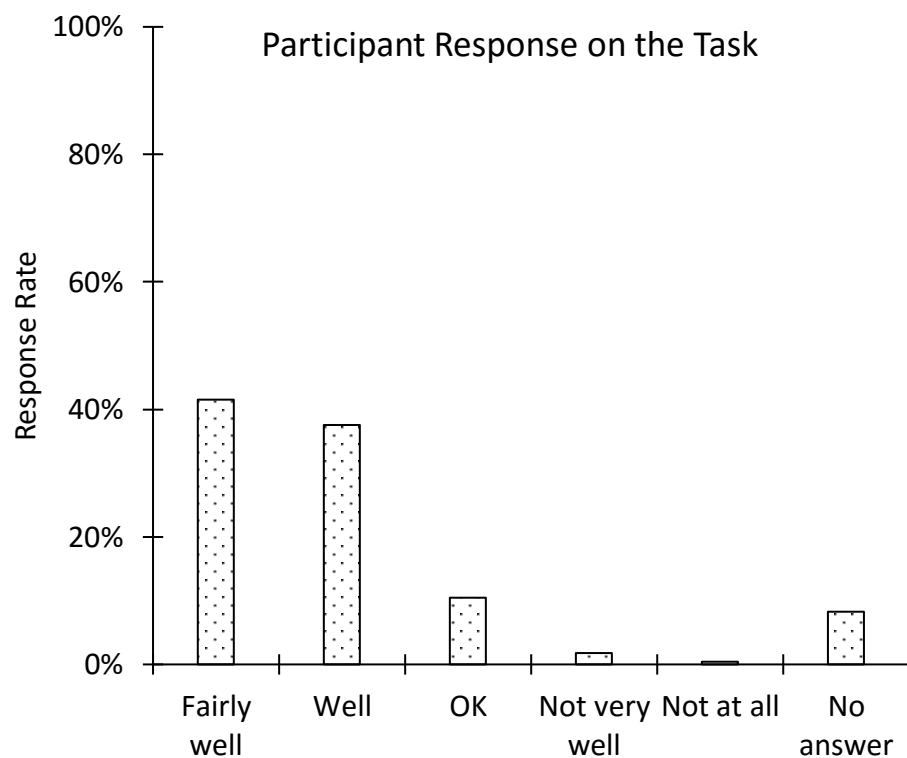


Figure 47. Participants' knowledge about the activity

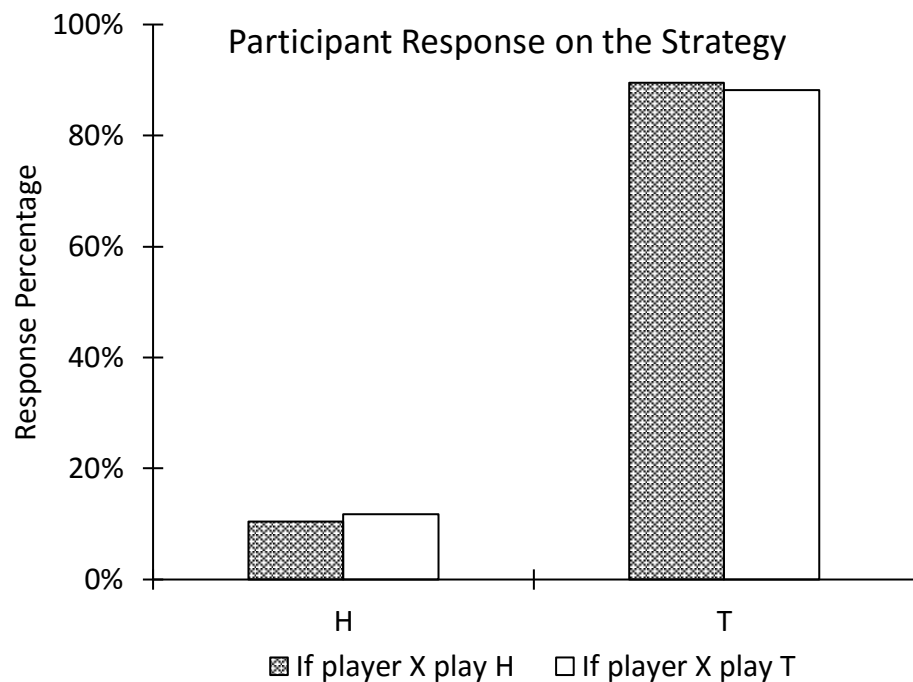
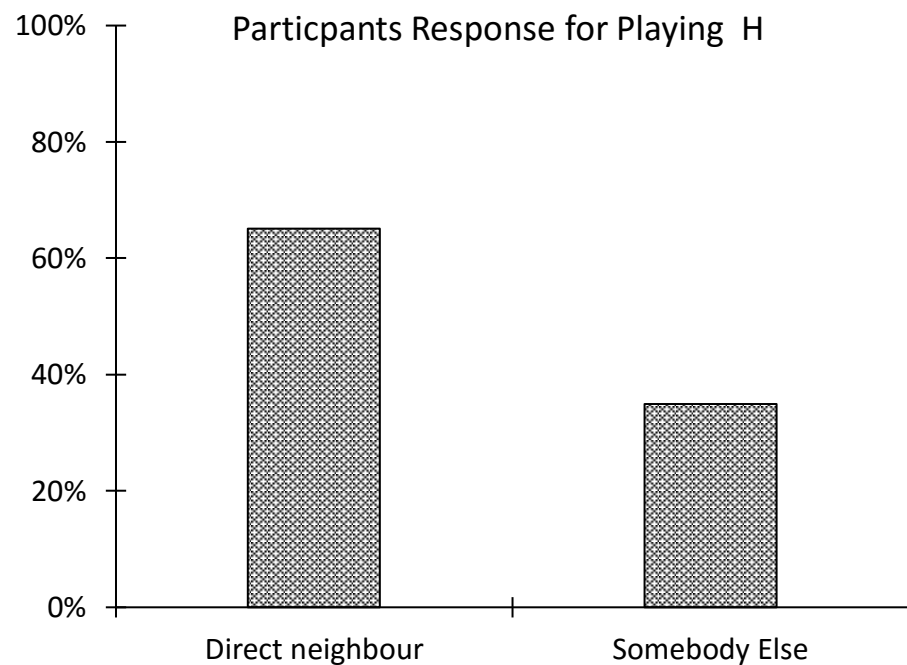


Figure 48. Participants' response against their opponent's choices

**Finding 1:** *The player Y will be more likely to play H with player X who is closer in proximity.* The previous studies (R1 and R2) established that the NE exists when individuals were required to make choices that could affect others. The behaviour was most profound when the individuals were required to punish another individual (Goddard et al., 2011; Goddard, 2012; Goddard et al., 2013; Noh et al., 2014). Hence, the main analysis for this study was to understand the extent of the NE in the context of a non-cooperative game, to determine whether the NE bias exists in a form of cooperation. In *Game 3* the participants were asked to choose how to play the CG with their direct-spatial neighbour(s) in a lateral (left/right) position and somebody else, which we identified as player *X* (refer to Appendix C1 to C6). The participants were required to make a choice as to who they play H with and who to play T. Figure 49 shows that 65.066% of 229 participants were more likely to play H with their direct-neighbour(s) instead of somebody else in the lecture theatre. From the

payoffs matrix in Figure 42, by playing H, the player indicated that, they would cooperate with their direct-neighbour(s) instead of somebody else, which shows NE in expression to avoid penalising their direct-neighbour(s).

Thus, we would like to explore if NE has significant differences if the direct-spatial neighbour(s) were individual in a lateral (left/right), longitudinal (front/back) and lateral and longitudinal (left/right and front/back) proximity.

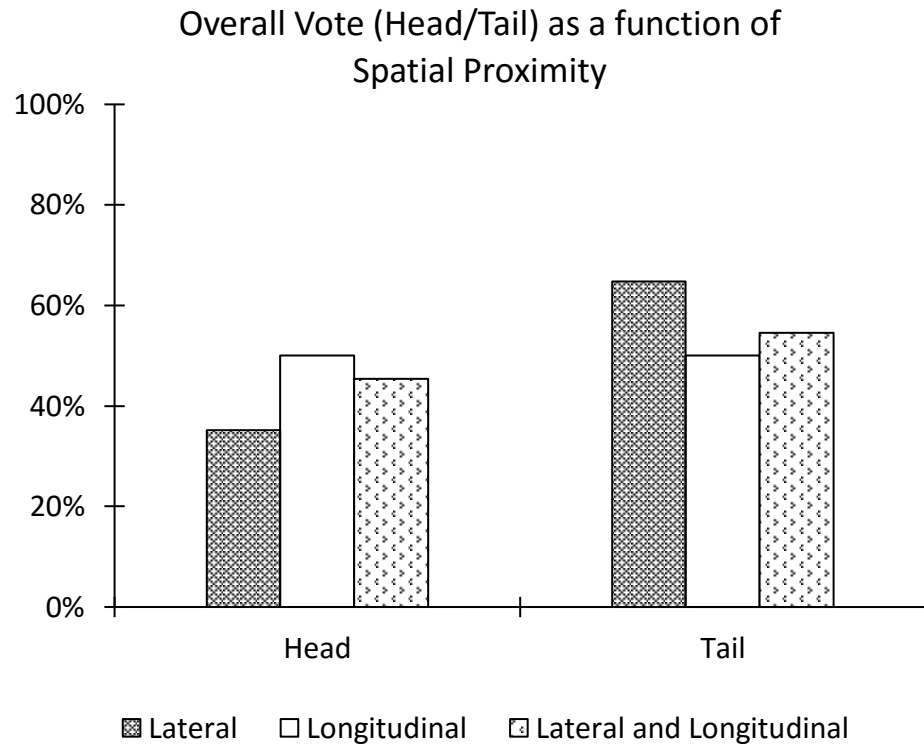


*Figure 49.* Response rate as a function of recipients' position. The response was made by the player *Y* when plays H against the player *X*.

***Finding 2: The player Y will be more likely to play T with player X who is closer in proximity with high visibility.*** The earlier analysis showed that the NE was most profound, when the neighbour(s) were at the lateral (left/right) position than somebody else in the lecture theatre (Finding 1). Which as the player, the participants were avoiding to defect their direct-spatial neighbour (s) by playing H. Hence, in consideration of *Game 1* and *Game 2*, the analysis aimed to distinguish if the proximity does affect NE, if the participants were given an option to play with participants either at lateral (left/right), longitudinal (front/back) position or both (lateral and longitudinal).

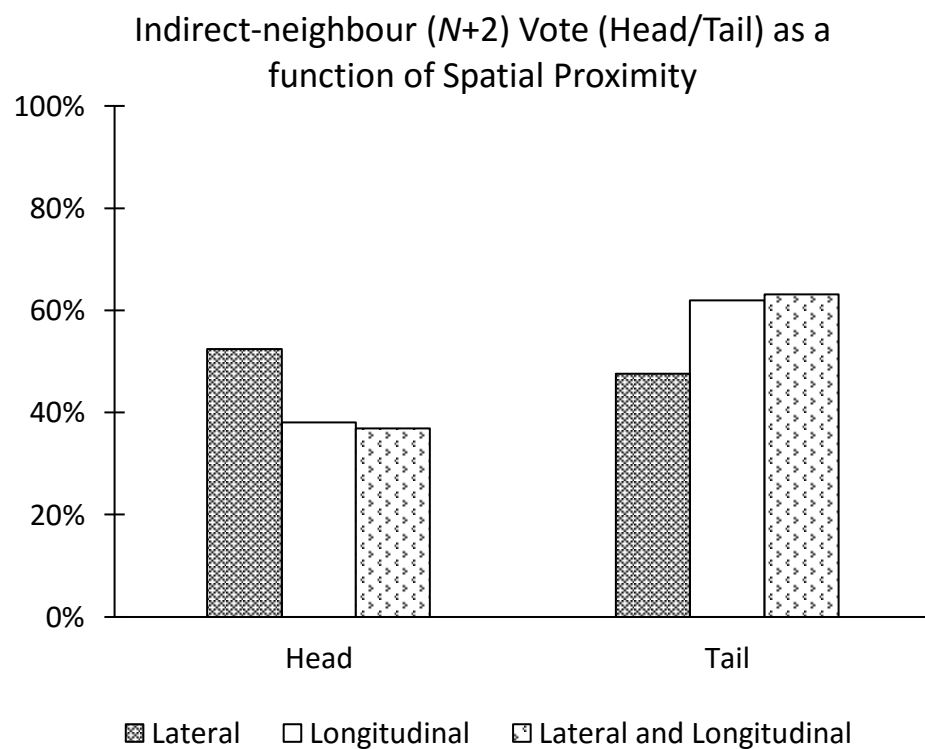
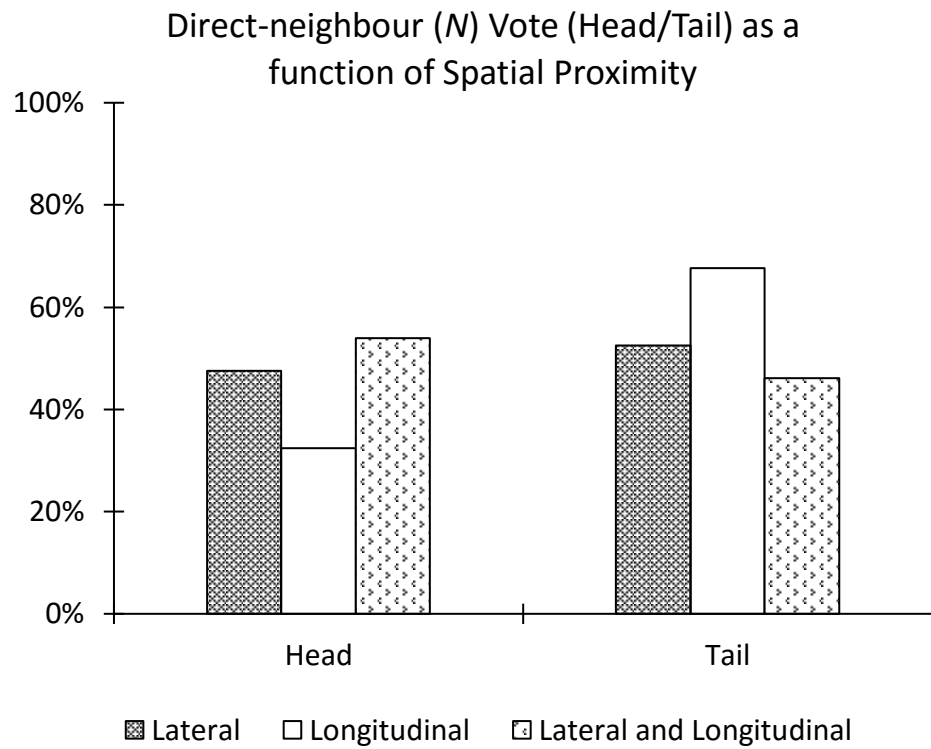
Figure 50 shows that participants were more likely to play T with player *X* at the lateral position in contrast to the longitudinal position ( $\chi^2 (1, n = 153) = 7.380, p = .007$ , small effect size, *Cramer's V* = 0.220). The analysis indicated that the participants were more likely willing to take a risk of defecting their lateral neighbour (s) rather than than their longitudinal neighbour (s).





*Figure 50.* Vote frequency as a function of spatial proximity when player *X* is at lateral, longitudinal and both (lateral and longitudinal) positions.

In Figure 51.1 (upper panel), the result shows that the participants were more likely to play T with their direct-neighbour(s) at lateral position (52.439%) and at longitudinal position (67.606%), compared to when their direct spatial neighbour(s) were at the both position (46.053%). And similar pattern was observed at longitudinal and both (lateral and longitudinal) position for indirect-neighbour (s) (see Figure 51.2, lower panel) for neighbour(s), in which the participants were more likely to play T with indirect-neighbour(s) two spaces away, 61.972 % and 63.158% respectively. However, when the participants were asked to play against neighbour(s) who were at the lateral position, the result show (Figure 51.2, lower panel) that participants were more likely to play H with the indirect-neighbour two spaces away, 52.439%.



*Figures 51.1 and 51.2. Vote frequency for as a function of spatial proximity. Figure 51.1 (upper panel) shows the direct-neighbour(s), while Figure 51.2 (lower panel) shows the indirect-neighbour(s).*

### R3.5 Discussion

This study was designed to test the *neighbour effect* (NE) in a context where there was likely to be a self-interest conflict against the observed spatial proximity bias. Participants had to choose either to cooperate or defect (play H or T) with other players that were their neighbour(s). The study showed significant NE when the participants had to pick another participant to cooperate with to gain a bigger payoff as a team, where they would likely to play **H** with their neighbour(s). This suggested that, the participants demonstrated a sympathetic equilibrium with the other player by choosing to cooperate even though, by playing **H**, the probability to lose is 50% (Sally, 2001). Approximately, 54.148% of the participants commented that their choices were either random or made by calculating the possible outcome that gave them a bigger payoff.

The NE is put forward as a bias where individual tends to avoid penalizing another individual. It is argued that, the differences in the physical settings (lateral/longitudinal positioning of individual) could affect the interaction order in which the proxemics bias might exist (Goffman, 1983; Hall, Birdwhistell, Bock, Bohannon, Diebold, Durbin,... Vayda, 1968).

The results show that, the NE was significantly associated with lateral (left/right) proximity more than longitudinal (front/back) proximity. Hence, it recommended that, the spatial proximity based on the positioning is vital to determine the level of cooperation with their neighbour(s). The participants were more associated with the NE when it was proximal side-by-side and closer in the seating arrangement against the position where

there was relatively low visual contact as in the longitudinal (front/back) position. In which contradicted the study done by Gardin et al. (1973) who suggested that there was a significant association between the level of interaction and level of visibility. However, for the proximal lateral condition it was always explicit which players were neighbours, but in the longitudinal conditions there was some ambiguity over the exact neighbour for players in the central block. This is unlikely to account for the difference between lateral and longitudinal NEs.

Thus, the significant association between NE and spatial proximity means that the null hypothesis that there would be no association between the *Voter*'s strategy and spatial proximity is rejected. Therefore, it is suggested that the *Voter* will be more likely to play ***H*** with player *X* who is closer in proximity with high visibility.

## **Chapter 4**

### **Discussion**

The **Discussion** section comprises four main parts (D1.2, D1.3, D1.4 and D1.5) which summarized the results from the three main studies, evaluation of the main contributions, discussion on the limitation and suggestions for future study.

## **D.1: General Discussion**

### **D1.1 Chapter Outline**

The starting point for this thesis was an observation of a proximity bias in the rather unlikely setting of a TV game-show. Thereafter the thesis aimed to verify the original observation and test its portability. This chapter gives a summary and evaluation of the main findings and contributions that can be gleaned. Limitations are discussed and speculations for future research proposed.

D.1.1 Chapter Outline

D.1.2 Summary of the Results

D.1.3 Evaluation of Main Contributions

D.1.4 Limitations

D.1.5 Suggestions for Future Study

D.1.6 Conclusion

## D1.2 Summary of the results

The two principal objectives were:-

1. To demonstrate/re-validate/replicate/extend the original observations of the *neighbour effect* using observations of the Weakest Link TV game-show to investigate the endogenous and exogenous factors in voting.  
and,
2. To test the validity and reliability of the observations from 1. by finding other scenarios where the *neighbour effect* can be tested.

The subsidiary objectives were conditional on 1. And 2.,

3. To test whether *neighbour effect* occurs with positive vote as well as negative vote (i.e., exogenous factors) in the Weakest Link
4. To assess *neighbour effect* on tasks involving strategy (e.g., Prisoner's Dilemma)

***Objective 1. To demonstrate/re-validate/replicate/extend the original observations of the neighbour effect using observations of the Weakest Link TV game-show to investigate the endogenous and exogenous factors in voting.***

Objective 1 was addressed in chapter **R1**. The original *neighbour effect* (NE) was *demonstrated/re-validated* and *replicated*. The original observations were also *extended* by introducing gender into the analysis. The original observations of Goddard et al. (2011) were one-dimensional by focusing only on proximity and not taking into account the

observations showing gender discrimination from studies by experimental economists. Once gender was introduced as a factor it was shown to make a significant impact on the contestants voting behaviour.

The original 72-episode study was reanalysed. This confirmed the original demonstration of the proximity/NE. A more substantial 151-episode study was required though because the reanalysis threw up a hitherto unexplained gender effect. The explanation for the gender effect remained obscure because the original study had failed to record contestants' accuracy fielding questions on the Weakest Link (WL) game-show. The replication resolved this issue by showing no difference in question accuracy between female and male contestants. The gender effect from the original study was not replicated but a different gender effect emerged that showed females and males were more likely to vote against males. This appeared to be taste-based discrimination because otherwise, at least based on their (male and female) ability to answer questions, there was no other obvious difference between genders. The source of the gender discrimination found in the original 72-episode study has not been resolved.

***Objective 2. To test the validity and reliability of the observations from 1. By finding other scenarios where the neighbour effect can be tested.***

If the interesting observation of the NE seen in the WL was to have any real significance, outside of the rather narrow academic community of WL scholars, then it needed to be shown as a bias that had a more universal impact on human social behaviour and decision-making. One aspect of WL that made it so special for researchers was that the particulars of the game format happened upon precisely the right conditions to expose



the psychological biases in voting that would ordinarily remain hidden. **R2** described the field experiment based on a lecture study voting activity. It was designed carefully to simulate some of the key features of WL that were postulated to nourish the NE. Chief among these key factors was uncertainty. Consensus studies from WL showed that once contestants become unsure whom to vote against then any bias, like NE or similar would be more likely to sprout. The uncertainty surrounding whom to vote for in the lecture study was massive because they were given no useful information as to whom to vote for. In the absence of any reliable game-specific situational information, participants had only their unreliable, dispositional hunches to go by. This latter source of information included the proximity bias responsible for the NE.

That the field experiment yielded such a compelling NE was a genuine surprise. Due to its scale, pilot studies were not possible beforehand and each field experiment was a major event in itself. Great effort was made to ensure that participants were unlikely to know other people in the lecture. This was the first major induction and orientation event that the students had been exposed to. Even if some early friendships had been forged and groups turned up together their random allocation to their unique seats made it unlikely that they would be seated by acquaintances. This random allocation went beyond the non-random contestants' configuration in WL. Although the study was designed to mimic the uncertainty in the low consensus WL voting rounds, any effect was predicted to be most likely masked by other factors. Also given the low stakes and relative insignificance of the task at hand to participants, made it seem unlikely that voting patterns would be anything other than noisy random picks. Certainly not the clear tri-phase pattern that mirrored the phenomenon from WL almost perfectly.

This observation of NE was probably the most important finding in the thesis. It meant that the proximity bias in WL was not merely a symptom of the TV game-show, rather it was more likely to be a universal facet of human social life, at least in the UK. Furthermore, embedded within the field experiment of the voting activity was the solution to the third objective, because now the NE could be created outside WL it could also be manipulated and probed.

***Objective 3. To test whether neighbour effect occurs with positive votes as well as negative votes (i.e., exogenous factors) in Weakest Link***

The field experiment of the voting activity produced potentially the second most important finding in the thesis as well as the most important. The possibility remained that the powerful NE recorded in the WL, and the field experiment, might be a fundamental bias associated with picking from any choice menu, irrespective of its type or valence. For example, consider what would happen if the negative vote made by WL contestants was replaced with a positive vote: voting for the SL rather than the WL. If the NE was a simple, generic, automatic voting effect then the same pattern would be seen, an NE and avoidance for picking the neighbour as SL. If, however, the NE was an avoidance for doing something "bad" to their direct-neighbours then switching the vote valence from "bad" to "good" should eliminate, or even reverse, the NE, as neighbours are chosen. This enabled a clear distinction to be made in the predicted outcome.

As it turned out, the change in vote valence from negative to positive eliminated the NE. It seemed that the proximity bias, the NE, was a reluctance to do something bad or

“nasty” to a neighbour. Whether this was borne out because of altruism or a fear of retribution was not resolved. The important thing about this observation though, is that the implicit, unconscious NE, was sensitive to the valence of the vote. It is almost as if there is an implicit ‘do no harm’ (to the neighbour) principle. This could form the foundation for a built in automatic bias for pro-social behaviour.

***Objective 4. To assess neighbour effect on tasks involving strategy (e.g., Prisoner’s Dilemma)***

The second subsidiary objective was to test whether the NE extended to judgments other than voting. This was assessed in **R3** detailing the second field experiment. The findings confirmed that participants played the PD game in a significantly different way with direct-neighbours rather than indirect-neighbours.

### D1.3 Evaluation of Main Contributions

The two main contributions from this thesis that deserve the attention of the wider academic community of researchers were the claim that this research has established the NE as a robust bias in decision-making and implicit social cognition and the advantages of the WL TV game-show as a research methodology in Psychology.

***Contribution 1: The neighbour effect as a proximity effect.*** The NE proved to be a consistent theme that ran through all of the studies in the thesis. It was shown to be robust and likely to be useful in understanding discrimination, implicit social cognition, altruism (pro-social behaviour) and cooperative (trust behaviour).

1. *Discrimination.* The NE in the WL emerged as a useful method for measuring in an indirect fashion implicit bias and discrimination. The results from **R1** confirmed that the WL exposed aspects of gender related discrimination largely in line with other US based studies of WL. Females and males tended to vote against males and/or for women.

2. *Implicit and Explicit information.* A wealth of research suggests that much of our social cognition, attitude formation, decision-making, judgements of attractiveness et cetera are mediated by unconscious, implicit processes. The NE exists as another key feature of our implicit social cognition. Prior to the research presented in this thesis there had been relatively little work on these proximity biases.

It was assumed that voting decisions were mediated by using two very different sources of information: *situational* (exogenous), the game-specific, public performance of the other players, and, *dispositional* (endogenous), their individual, internal, subjective-dependent attributions. In rounds where *situational* information was unequivocal, so the weakest player was easily identified by the other players (high consensus), there was no voting bias. However, significant biases emerged as uncertainty increased (consensus decreased) about the identity of the weakest player. In the absence of clear-cut *situational* information, because all players performed equally well (or badly), players resorted to their private, bias-prone *dispositional* information source. A third information source was also likely to be available that referred to the contestants' ability to react to other contestants' voting decisions. This source of information was likely to be conscious and deliberate. This third source emerged in the literature review (Antonovics et al, 2005; Levitt, 2004) and through the voting biases in the WL second-round, reciprocity and consistency. Although, this strategic game-playing was treated as a third source of information bias, further research is required to determine if it is a separate TYPE III bias, or whether it could just as easily fit with the TYPE I, situational source described above in the form of information-based discrimination.

3. *Altruism and Fairness.* The sensitivity of the NE that was seen once the vote valence switched suggests that the NE might serve some function as an implicit bias favouring their closest neighbour(s) more than participants who were further from them as an act of altruism (Baron, 1995; Bell, 1991). It remains to be seen whether the altruism to the neighbour is a genuine selflessness, or a cooperative strategy to avoid conflict.

The retaliation witnessed in second-round voting suggests that there was some good reason to expect retribution. If conflict was an inevitable risk of doing something bad to somebody, then it would be reasonable to assume that the feud should be initiated with a contestant further away than the direct-neighbour. Altruism might also have mitigated the extent of second-round consistency voting. Although on the face of it, consistent voters were significantly more likely to vote for same contestant again in second-round, once the voting of controls was taken into consideration, it actually showed that the consistent voters were significantly less consistent than controls. Their reluctance to target the same votee again in the second-round, even though justified by the controls, is a kind of altruism that might reflect their reluctance for conflict. These hypotheses require further elaboration and testing.

4. *Cooperation and Trust.* The NE was strongly dominant when the contestants were favouring their closest fellow contestant(s) (i.e., direct-neighbour) in contrast to contestants that were further from them. Proximity played an important role to explain this behaviour when they were making the decision with regards to any social context (i.e., WL observation, voting activity and the coin's game activity). This supported construal level theory where spatial proximity effects perception and decision making (Trope & Liberman, 2010). Additionally, it was also suggested that the NE was more obvious at the lateral (side-by-side) proximity than the longitudinal (front and back) proximity.

The NE explained the level of cooperation and trust shown by participants in the second field experiment study (PD). PD is a dilemma for the player as they must take into consideration the likely strategy of the other player. Participants were more likely to play

H (cooperate), with the presumption that this was the optimal strategy when the participants believed that the other player would play H. The implicit NE worked twofold: to influence the inclination to cooperate with a neighbour(s) and influencing the implicit assumption that the neighbour will most likely act cooperatively too. When players play against direct-neighbours they are more likely to pursue a pareto-dominated strategy even though the pure strategy to defect, according to the Nash equilibrium, should dominate when a rational player seeks to maximize payoffs.

5. *Proximity.* The observations of WL and the set-up of the field experiments were designed in such a fashion so as to create the optimal conditions for creating the kind of proximity bias associated with the NE. These conditions have been detailed previously but mimic the uncertainty associated with low consensus voting rounds in WL. The strength of the NE in both the observational field studies of WL and the field experiments was nevertheless a surprise. The consistency of the pattern of voting as a function of voter-votee spatial relationships was also a surprise. It typified a three phase pattern. In the cases highlighted where the vote carried a negative valence, the proximal phase was characterized by a neighbour avoidance effect. This was followed by a medial phase, where the contestants located most centrally received more votes than expected. Finally, the distal phase had the contestants furthest away voted for as expected.

Construal Level theory (CLT) would suggest that the direct-neighbours were most concrete in the voter's thinking and benefit accordingly from their benevolence, the furthest become the most abstract and so are largely disregarded whereas those located centrally fall between the extremes and suffer as a consequence of the charity afforded the

neighbours. The suggestion is that the central contestants were not targeted because of their centrality but because voters tend to avoid votes for neighbours.

Another related explanation for the tri-phasic proximity effect was that it was the manifestation of two choice biases. The first choice bias was the oft discussed NE. This was followed by the second choice bias for centrality preferences among similar options. Shaw, Bergen, Brown & Gallagher (2000) showed that participants have a strong automatic and unconscious bias for selecting the middle one of three items in an array. Taken together these biases combined would lead to the pattern of performance exhibited by contestants on the WL and on the fresher in a lecture.

The CLT provides a psychological explanation whereas NE allied with centrality bias is based more on choice theory. It could even be the case that the NE and centrality biases become the manifestation of CLT.

***Contribution 2: The TV Game-show is a Useful Tool to Observe Behaviour.*** The review of the WL game-show as a methodology in **I2**, showed that a disparate set of researchers from different academic fields have converged on the utility of using WL as an empirical testing bed for their theories. In doing this a WL contestant was considered variously as a potential source, or target, of discrimination (explicitly or/and implicitly), as a rational and strategic game player and as a decision maker prone to unconscious bias.

*The WL contestant as prejudiced, strategic, biased, decision maker.* Key features of the vote were that it involved an eight alternative-forced-choice (8AFC), it was cast



under effectively blind conditions without knowledge of the votes of other contestants, it was privately made.

Putting oneself in the proverbial shoes of the contestant, it is reasonable to assume that during the first-round of questions that they would be in a state of heightened anxiety and excitement due to their unusual social context. They were appearing in public, in an unfamiliar situation on a TV game-show that was scheduled to be broadcast nationally. They were facing a set of unknown questions to be answered quickly and without error and faced the prospect of elimination should they perform badly. Taken together it is reasonable to assume that their decision-making would be affected by these external situational factors many beyond their control. Moreover, the structure of the game-show and the atmosphere created by the host were designed to make the contestants uncomfortable and ill at ease. On top of all of this they were put in the very uncomfortable position of having to pick someone, a relative stranger, to do something bad to. Taken together these different factors mean that their performance demands their upmost attention and makes the WL an ideal methodology.

The WL confers many advantages as a research tool. Perhaps, the main advantage the WL offers is as an alternative method for studying sensitive topics such as race and gender discrimination. For example, it is difficult for experimental economists to examine factors such as discrimination in active labour markets. In the laboratory, participants are also unlikely to reveal hostile discriminatory behaviour. In Psychology, it is also often difficult to measure such behaviour systematically, either in a laboratory setting or through observation elsewhere due to ethical considerations. However, in the WL, it was easy to

study specific demographic variables of contestants, like race and gender, and relate the measured variables through to their game playing behaviour. In this way gender/race could be investigated in the WL in an unobtrusive fashion that was not likely to inhibit or even be noticed by the participants. In other words, the participants were unaware that their behaviour was under scrutiny.

In a similar fashion, the WL offered a unique way to study the voting behaviour of contestants. In particular, a vote can be seen as choosing to do something “bad” (i.e., voting out from the game) to another contestant. It is difficult to devise scenarios to entice people to do ‘bad’ things in the laboratory that satisfy both ethical guidelines and have a meaningful credibility for participants.

Taking these factors together means that the WL offered a perfect indirect method for studying implicit social cognition. The format of the game itself and its popularity afforded the possibility of analysing a very large data set. For instance, in this thesis, 2,583 votes were observed. A figure that is not feasible to replicate by recruiting participants in laboratory based studies. Another advantage of the game-show format is that the WL game-show offered a high degree of effective control between episodes, which is something that is almost impossible to be observed in a normal laboratory setting.

The information about the WL game-show was accessible and almost as visible to researchers as it was to the contestants. Whereas, there was still the possibility that researchers were not privy to off-screen interactions between contestants, there was also the advantage that the researcher could replay and reanalyse the game over and over again,

something that a contestant was unable to do in real time. The contestants taking part were fully familiar with the game-show and the rules and so were unlikely to be susceptible to the kind of *participant effects* that often blight experimental work, where participants try to anticipate the goals of the research in tasks that can appear confusing and contrived. The WL also allowed the researchers to use independent observers to code the contestants' performance reducing the possibility of introducing experimenter bias. The task was simple, countable and observable meaning it was unlikely that experimenters could introduce subjective interpretations particularly as data from WL was collated and corroborated by independent coders.

In summary, the WL provided an empirical test-bed for theories in social psychology, decision-making and economics. The WL as a method was effectively double blind by being relatively free from participant effects and experimenter bias. The task was credible and authentic for participants because the goal made sense to them. Their choices were natural and last, but not least, the high stakes offered in the game based TV game-show is something that is impossible to be executed or replicated in a laboratory setting.

However, there were also important limitations to studying WL. Despite the large sample size, it was not representative the population in general. The sample was by its very nature self-selected, comprised only of the kind of participants that put themselves forward to appear in TV game-shows. Furthermore, it was selected further by the producers that presumably selected on the basis of the kinds of attributes that would be deemed attractive to a TV audience. There was a lack of researcher control over the participants and their briefing and arrangement in the game.

Another limitation was the contrived context where participants made their choices. Being filmed for a TV game-show is not necessarily typical of *normal* day to day behaviour. Although this unusual context could also be considered as an advantage as it required participants to take the task seriously. The laboratory cannot easily recreate a context where participants' decisions are open to such public scrutiny.

## D1.4 Limitations of the Study

In spite of the findings above there were some limitations to be addressed in this thesis.

**Limitation 1: Difficulty isolating *neighbour effect*.** It was documented above that the clarity of the tri-phasic proximity effect coming out of the field experiment voting activity was a surprise. This was largely because in many voting decisions or choice scenarios, there are many factors that can make up or dominate a choice decision. In WL, a contestant's primary choice information was how players play the game. Only in the absence of this kind of information does the proximity bias come to the fore. This made it very difficult to find instances of the NE in experimental settings. Aside from the field experiments, most of the laboratory studies piloted failed to show any consistent results.

Several pilots (refer to **M1.3**) were run in an attempt to replicate the findings from WL regarding the NE. These attempts centred on recruiting small groups of students to work on an initial group task that later required each of them to "pick" someone from the group on some pretext. The idea was that once the participants got around to picking one of their peers then they would display the NE. Several games, and their variants, were used to entice some kind of choice from the participants. The tasks were in the form of popular party or parlour games like *wink murder*; angels and detectives and even variants of the WL. However, no consistent NE findings emerged from any of these pilot investigations that typically turned out to be fraught with practical difficulties.

Even if the students were not close friends they frequently were acquaintances of each other which was likely to affect the choices they made. Recruiting the student participants and setting up a game playing scenario took, including briefings on playing the game, upwards of one hour and beyond, following which seven bits of data were recorded which were the voting decisions of the seven participants. One way round this was to repeat several rounds of the game but then the voting took on features of retaliation and consistency voting that typified the second-round voting pattern on WL. The only consistently reliable experiment used the field experiment lecture based studies.

**Limitation 2: Dynamic interaction.** This thesis was restricted to fixed interaction is in a static social context. For instance, in the WL TV game-show as well as the field experiments, all the participants were positioned at one point without having to move to another point. In the context of daily life, events are experienced as a temporal series of dynamic interactions, through which the spatial relationships between people is usually changing rapidly as they move from one point to another. For example, in road traffic congestion, or even when people are walking on the pavement. Hence, it will be interesting to test whether the NE is still a robust bias in dynamic social interactions. Testing this remains a challenge.

**Limitation 3: Cultural differences.** Another potential limitation of this study was an absence of cultural differences as all the participants were predominantly UK residents. As reported by Hall (1963), different cultures have different interpretations of closeness “...it became apparent that people from different cultures interacting with each other could not be counted on to attach identical meanings to the same or similar measured distances between them” (Hall, 1963, p. 1003). Hence, sensitivity to proximity varies according to social norms and it remains to be established whether the NE is a universal human trait or a cultural norm.

## D1.5 Suggestions for Future study

**Community Impact.** Nieuwenhuis, Völker and Flap (2013) surveyed residents' relationship in the Netherlands. They suggested that the likelihood of a negative relationship in the neighbourhood was associated with the likelihood that the residents had rented their property. The likelihood that they wanted to leave the neighbourhood was also related to the mode of property ownership. One argument that they put forward was that the healthiest neighbour/neighbourhood relations were realized by citizens that had cast a long shadow into the past and looked to cast a long shadow into the future as long term residents. It is likely the strength of the NE in these different kinds of communities would also shift. One area where the implications of NE can be investigated is within real communities with meaningful neighbours. Following the lead of Nieuwenhuis, Völker and Flap (2013) the concept of neighbourliness can be developed and potentially measured in communities.

**Tests of *Neighbour Effect*.** Tests of NE like those in WL and the field experiments provide an indirect measure of implicit social bias. Although the pilot attempts to generate NE failed, Valenzuela and Raghubir (2007) found that a replication of WL using 67 students demonstrated authentic game-playing scenarios. They managed this using a clever variant of WL that involved footage from the original episodes and included the students as an extra player in WL. The students were required to choose which player from the actual WL they would prefer to face in the final. They introduced a twist however by putting students in a cooperative or competitive condition. The cooperative condition meant that the student should imagine that they were playing the game



cooperatively and would share the eventual prize fund. In the competitive condition, however, the student had to choose whom to face under the usual rules where there was only one victor. In the cases when the students were allocated the competitive condition, 59% elected to play against a female contestant. In the cooperative condition only 21% of students chose to play with a woman finalist. It seemed that the students would prefer to face a woman in a competitive final but would rather play with a man if the proceeds were to be split. They go on to suggest that one reason why women were retained in the early rounds of WL was because they were perceived to be less threatening than men, possibly accounting for the positive discrimination towards women contestants seen in the WL in **R1** and in Antonovics et al., 2005).

**Self-Reflection.** Voting behaviour was used as an unorthodox alternative to traditional methods in psychology to try and explore how people make social choices in their everyday lives. It was challenging at first using participants that were contestants on TV shows but it became very exciting to find out that the behaviour that they showed on TV was found as predicted in the studies that I had designed. Therefore, this research established an alternative method to investigate individual choices, particularly relating to spatial proximity.

The study found that the Neighbour Effect is an important phenomenon, a spatial bias in decision-making that was hitherto largely unknown until I confirmed it in my studies. From a personal point of view I feel privileged to have happened upon a facet of human nature that was unknown before my work. The real challenge is to develop these discoveries for publication and wider dissemination.

## D1.6 Conclusion

“A bad neighbour is as great a plague as a good one is a great blessing”

(Nieuwenhuis. Völker & Flap, 2013, p. 1)

A robust proximity bias was demonstrated that affected voting decisions. The source of this bias was unconscious and automatic. The bias worked by predominantly favouring direct spatial neighbours. The *neighbour effect* was an avoidance to vote against neighbours if the consequence of the vote was perceived as being negative for the recipient. When the consequence of the vote was perceived as positive the *neighbour effect* disappeared. Therefore, the *neighbour effect* was not just a simple bias of choosing, as it changed according to the perceived impact the vote choice had on the neighbour. The analysis of TV game-shows served as an important indirect method to test theories in Psychology, discrimination and decision-making to investigate the endogenous and exogenous factors in voting behaviour.

## **Appendices**

The **Appendices** section comprises the materials used in the lecture studies for 2013/2014 (Appendix A), 2014/2015 (Appendix C) and 2015/2016 (Appendix B) cohorts.

## Appendix A

### SHEET 1 Ethics, Background and Instructions

#### BACKGROUND AND INSTRUCTIONS

1. **Ethics - 'Credit points' and 'Lottery'**:- Any research that gets carried out in the University has to be approved by the relevant ethics committee. I have to ensure that everybody here:-
- i.) Is fully informed as to what the procedures are in advance.
  - ii.) Has the right to confidentiality/anonymity.
  - iii.) Gives their informed consent to take part.
  - iv.) Has the right to withdraw, without losing rewards/benefits.

Everybody here is entitled to **one** 'Credit Point'\* and everybody here will be entitled to a minimum of **one** ticket to a lottery\*. (Some people might get more than one ticket but nobody will get less than one.) The lottery prize will be.....

2. **Background**:- We have carried out a bit of research already that has looked at 'voting' behaviour where:-

- I.) 'voters' have had '*open*' information about whom to vote for and why,
- II.) and, the vote was '*open*' (the 'voter' announced their 'vote' publicly),
- III.) and, the vote was *POSITIVE* or *NEGATIVE*.

3. **Aim for Today**:- We want to collect some Control/Normative data today to use as a baseline for what we've done before, described in 1.) above, but

- I.) 'voters' have '*closed*' no information about whom to vote for and why, and
- II.) the 'vote' is *closed* (the 'voter' makes their vote privately).
- III.) and the vote will be *POSITIVE*, *NEUTRAL*, *NEGATIVE*\*\*.

4. **Procedure**:- We will ask you to make **THREE** votes that vary according to **spatial scale**.

- I.) Vote for someone in the **LECTURE THEATRE** by making a cross on the seating plan.
- II.) Vote for someone in your **BLOCK** by making a cross.
- III.) Vote for someone in your **BLOCK** and in your **ROW** by making a cross.
- VI.) MOST IMPORTANTLY, please indicate your seat number here and make a note of it:-

I AM SAT IN BLOCK:-

I AM SAT IN ROW:-

I AM SAT IN SEAT:-

#### RULES

**Please be quiet until everyone has made their votes and the experimenter**  
**The only rule is that you can't vote for yourself!!!!**

5. **Thank you!!!!** Please make your judgements on the attached sheets and when you have finished just place your answer sheets in the envelope in which they came.

### About YOU!

M/F

Age

## Appendix A1

### Your vote is **NEGATIVE** and removes **ONE** voucher

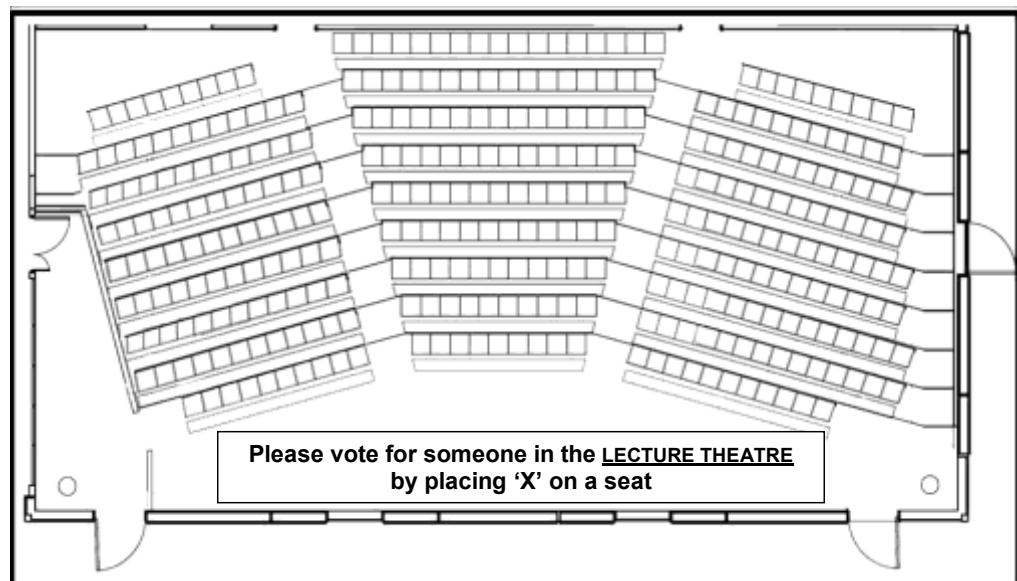
#### **About YOU!**

We would like to know a little bit about you so we can describe the demographics of our sample (please leave blank if you do not wish to disclose your gender/age). If you do not want to consent to take part in the study then simply do not make any votes. Even if you don't make any votes though, you will still gain a 'Credit Point' and voucher for the raffle so please put in your seat number. When you have finished please place this sheet back in the envelope provided. The whole process should only take a couple of minutes. There are no 'correct' answers and the only real guidance is that you can not vote for yourself.

I am female/male and I am \_\_\_\_\_ years old.

I am in Seat number:-

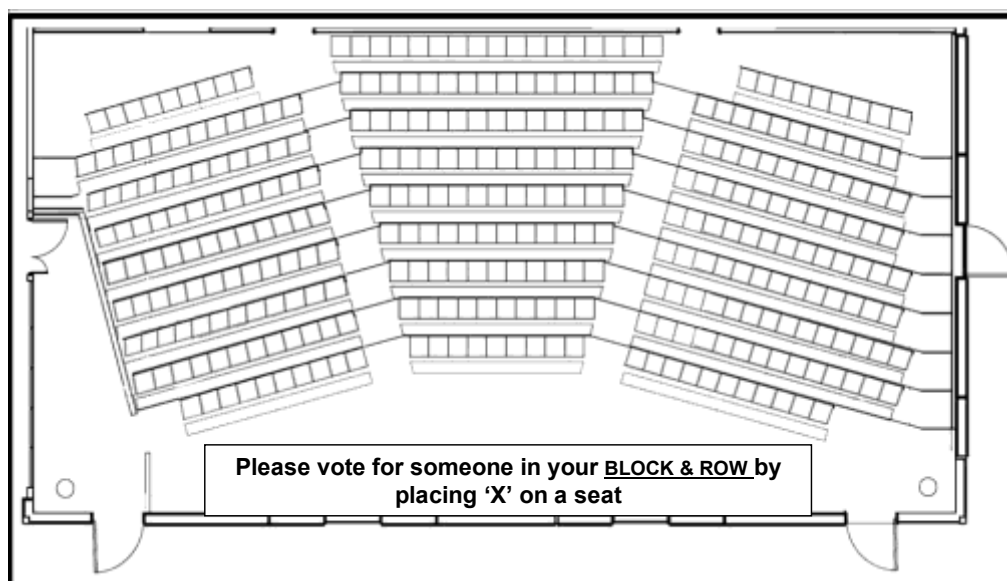
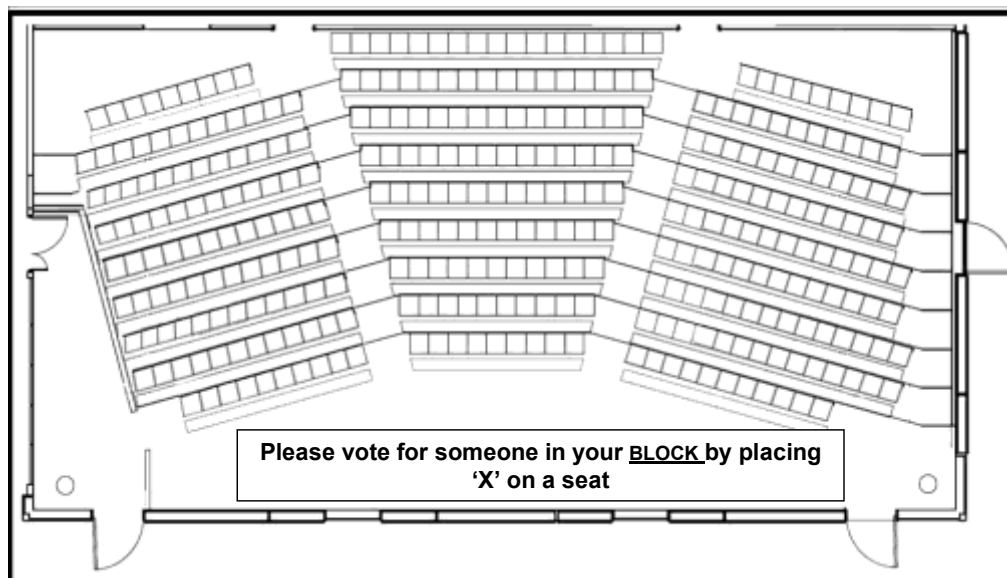
**IMPORTANT: Keep a personal record of your seat number!!**



**PTO**

## Appendix A2

Your vote is **NEGATIVE** and removes **ONE** voucher



This final question is optional and open-ended about how you made your choice. I made my choices by:-

.....  
 .....  
 .....  
 .....(optional)

**THANK YOU! Paul, Patrick, Adrian, Zamira, Chris.**

## Appendix A3

### Your vote is **NEGATIVE** and removes **FIVE** vouchers

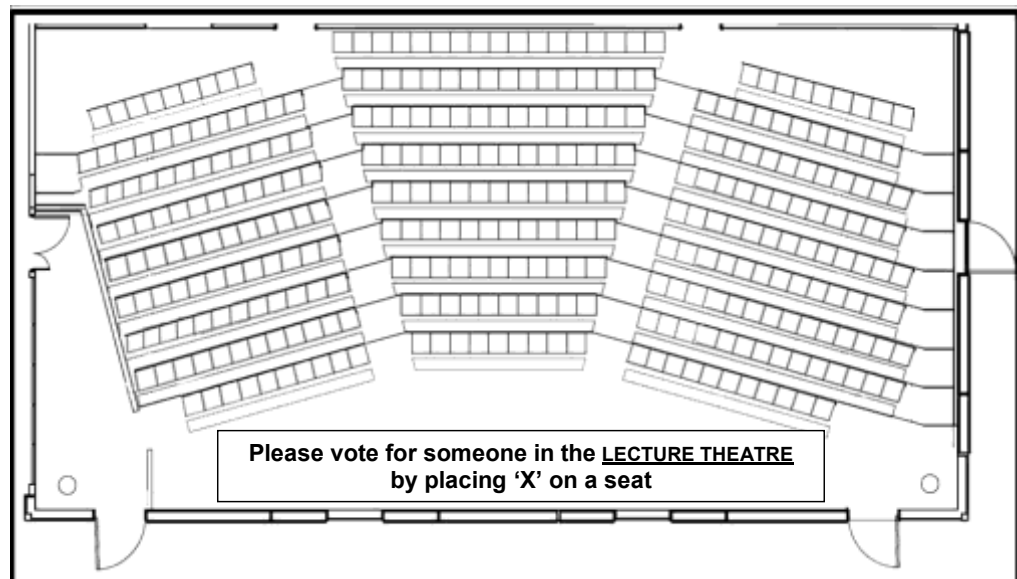
#### **About YOU!**

We would like to know a little bit about you so we can describe the demographics of our sample (please leave blank if you do not wish to disclose your gender/age). If you do not want to consent to take part in the study then simply do not make any votes. Even if you don't make any votes though, you will still gain a 'Credit Point' and voucher for the raffle so please put in your seat number. When you have finished please place this sheet back in the envelope provided. The whole process should only take a couple of minutes. There are no 'correct' answers and the only real guidance is that you can not vote for yourself.

I am female/male and I am \_\_\_\_\_ years old.

I am in Seat number:-

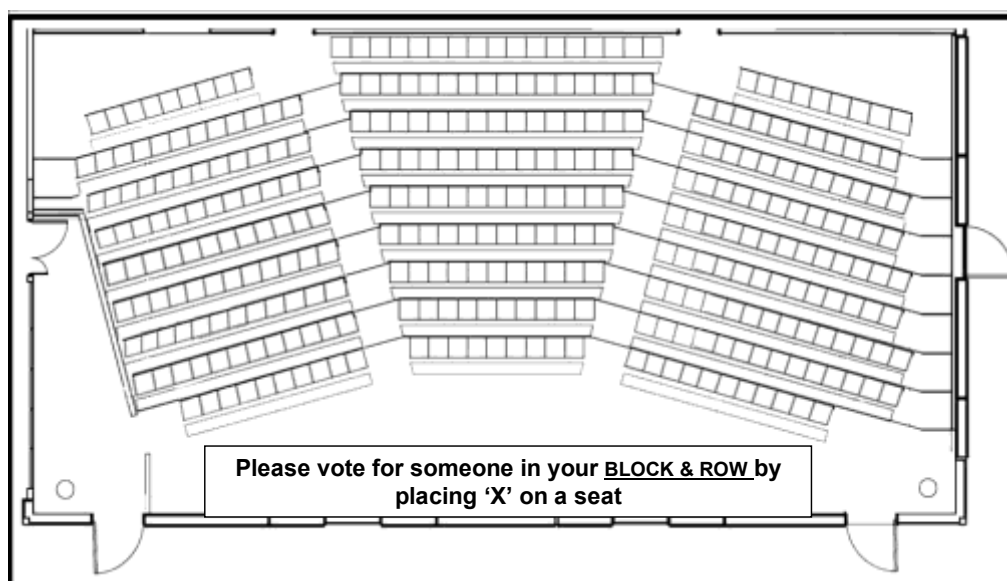
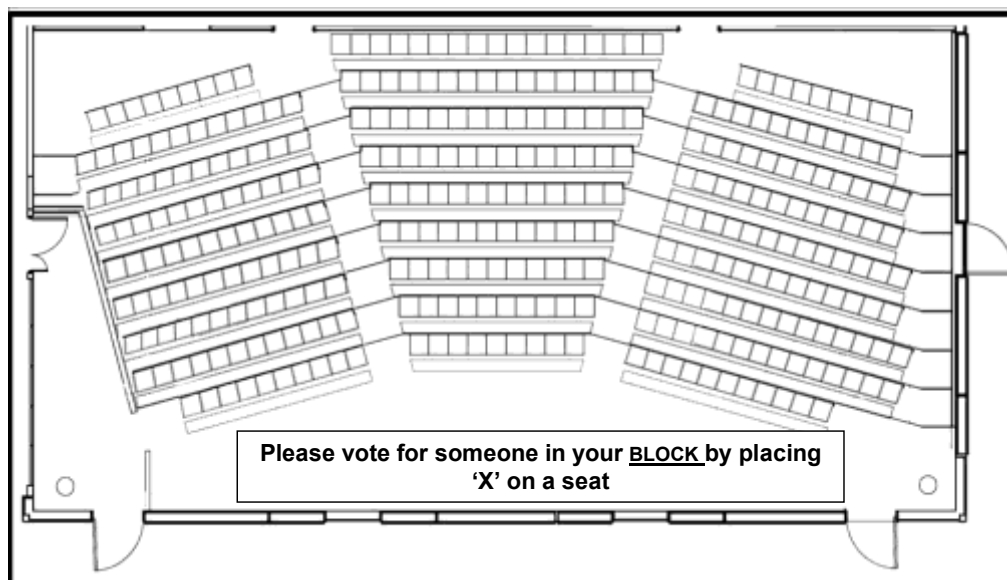
**IMPORTANT: Keep a personal record of your seat number!!**



**PTO**

## Appendix A4

Your vote is **NEGATIVE** and removes **FIVE** vouchers



This final question is optional and open-ended about how you made your choice. I made my choices by:-

.....  
 .....  
 .....  
 .....  
 .....(optional)

**THANK YOU! Paul, Patrick, Adrian, Zamira, Chris.**



## Appendix A5

### Your vote is POSITIVE and adds ONE voucher

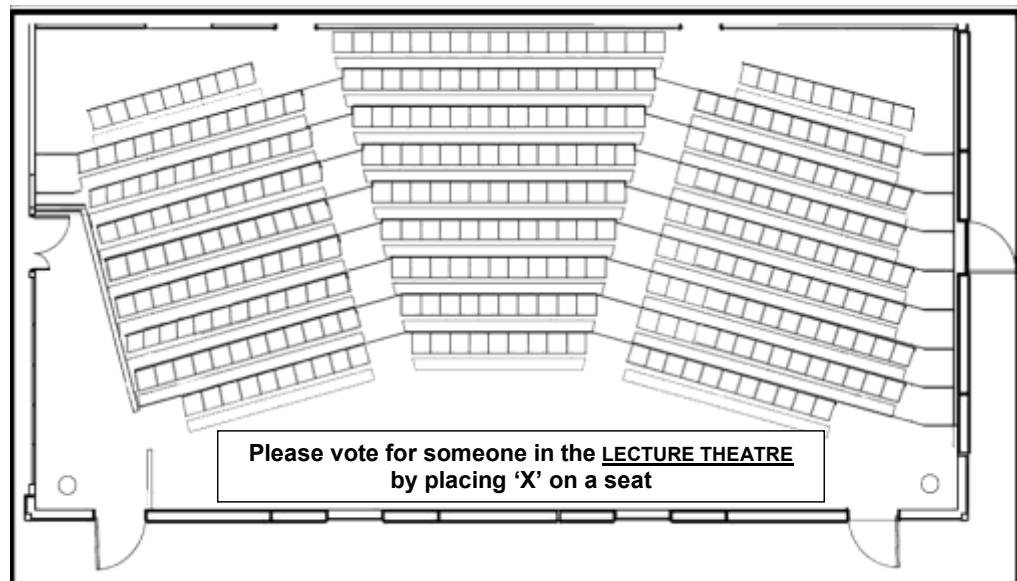
#### About YOU!

We would like to know a little bit about you so we can describe the demographics of our sample (please leave blank if you do not wish to disclose your gender/age). If you do not want to consent to take part in the study then simply do not make any votes. Even if you don't make any votes though, you will still gain a 'Credit Point' and voucher for the raffle so please put in your seat number. When you have finished please place this sheet back in the envelope provided. The whole process should only take a couple of minutes. There are no 'correct' answers and the only real guidance is that you can not vote for yourself.

I am female/male and I am \_\_\_\_\_ years old.

I am in Seat number:-

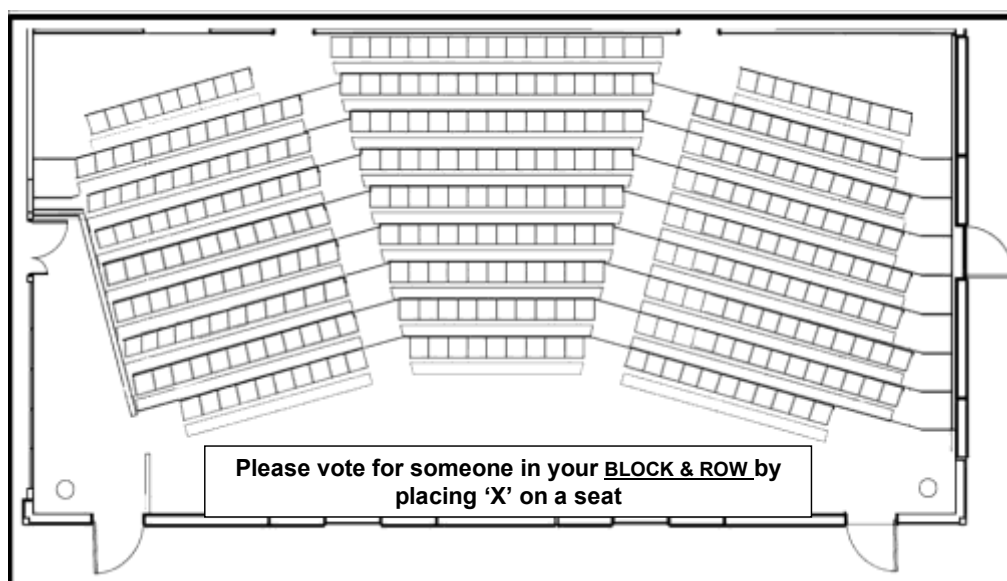
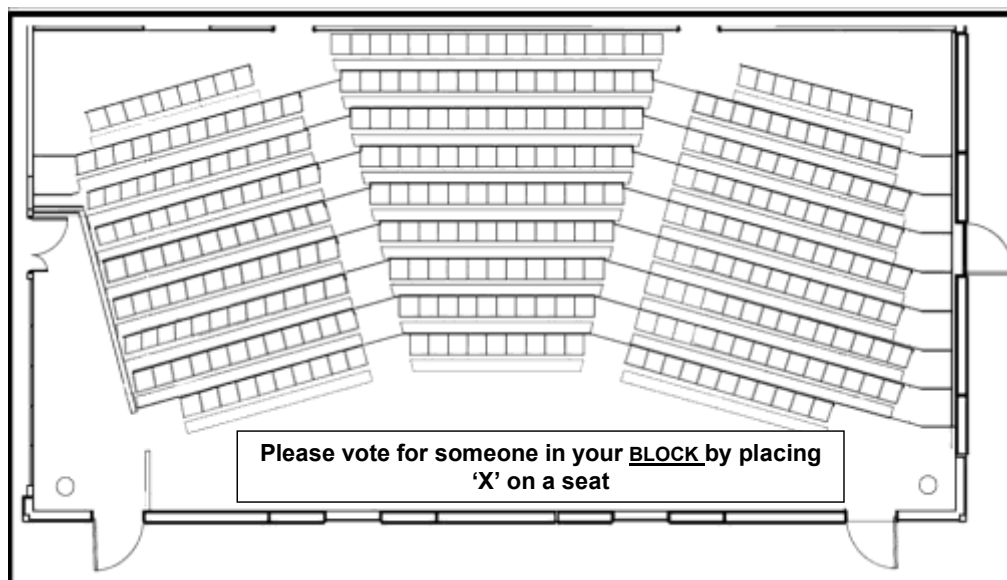
**IMPORTANT: Keep a personal record of your seat number!!**



**PTO**

## Appendix A6

Your vote is POSITIVE and adds ONE voucher



This final question is optional and open-ended about how you made your choice. I made my choices by:-

.....  
 .....  
 .....  
 .....  
 .....(optional)

THANK YOU! Paul, Patrick, Adrian, Zamira, Chris.

## Appendix A7

### Your vote is POSITIVE and adds FIVE vouchers

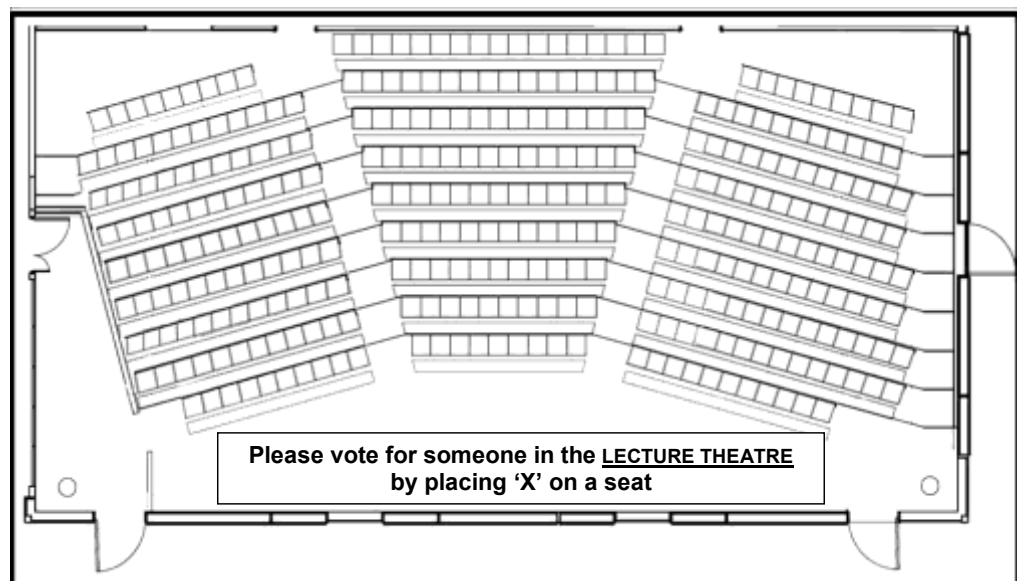
#### About YOU!

We would like to know a little bit about you so we can describe the demographics of our sample (please leave blank if you do not wish to disclose your gender/age). If you do not want to consent to take part in the study then simply do not make any votes. Even if you don't make any votes though, you will still gain a 'Credit Point' and voucher for the raffle so please put in your seat number. When you have finished please place this sheet back in the envelope provided. The whole process should only take a couple of minutes. There are no 'correct' answers and the only real guidance is that you can not vote for yourself.

I am female/male and I am \_\_\_\_\_ years old.

I am in Seat number:-

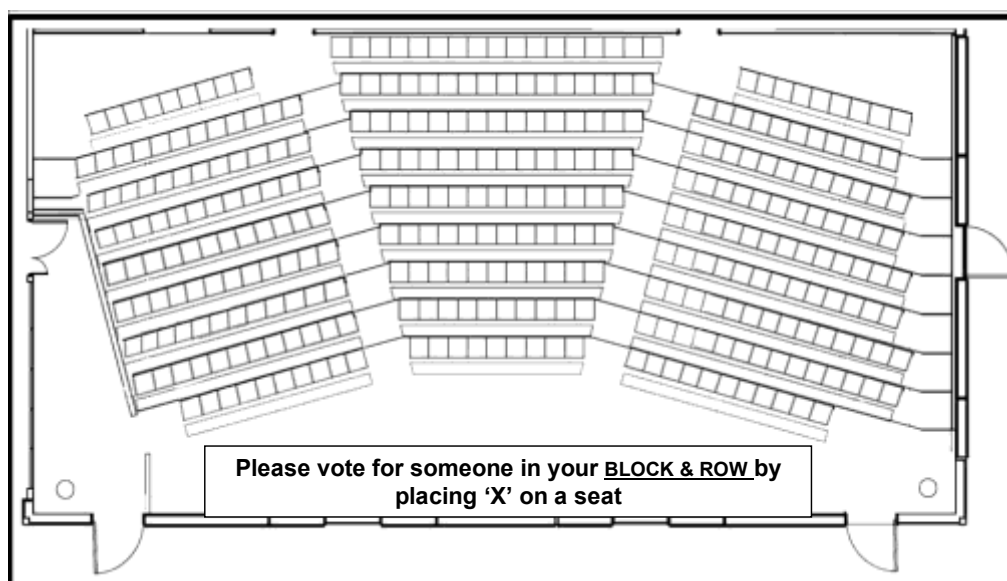
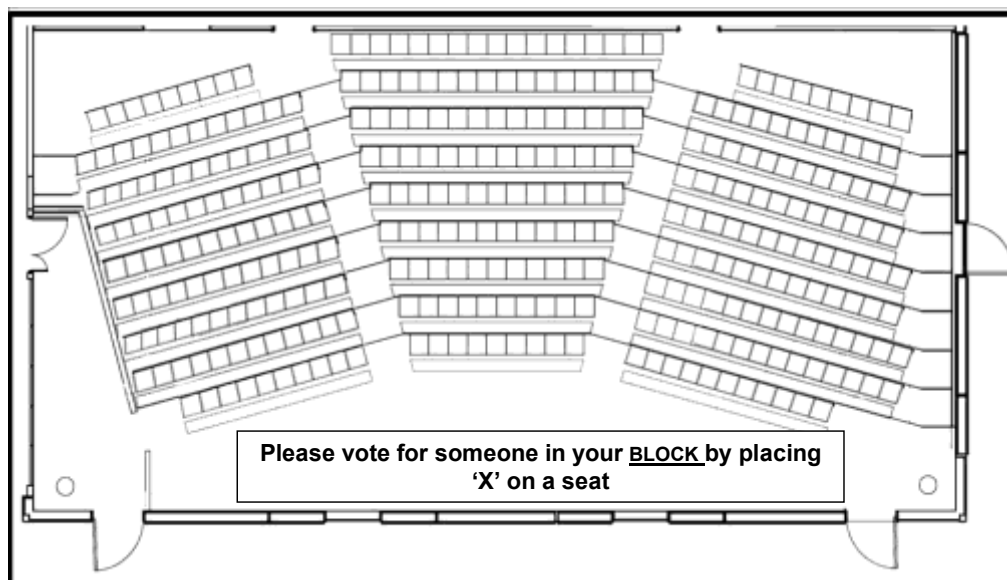
**IMPORTANT: Keep a personal record of your seat number!!**



**PTO**

## Appendix A8

Your vote is POSITIVE and adds FIVE vouchers



This final question is optional and open-ended about how you made your choice. I made my choices by:-

.....  
 .....  
 .....  
 .....  
 .....(optional)

THANK YOU! Paul, Patrick, Adrian, Zamira, Chris.

## Appendix A9

### Your vote is NEUTRAL and gives ZERO voucher

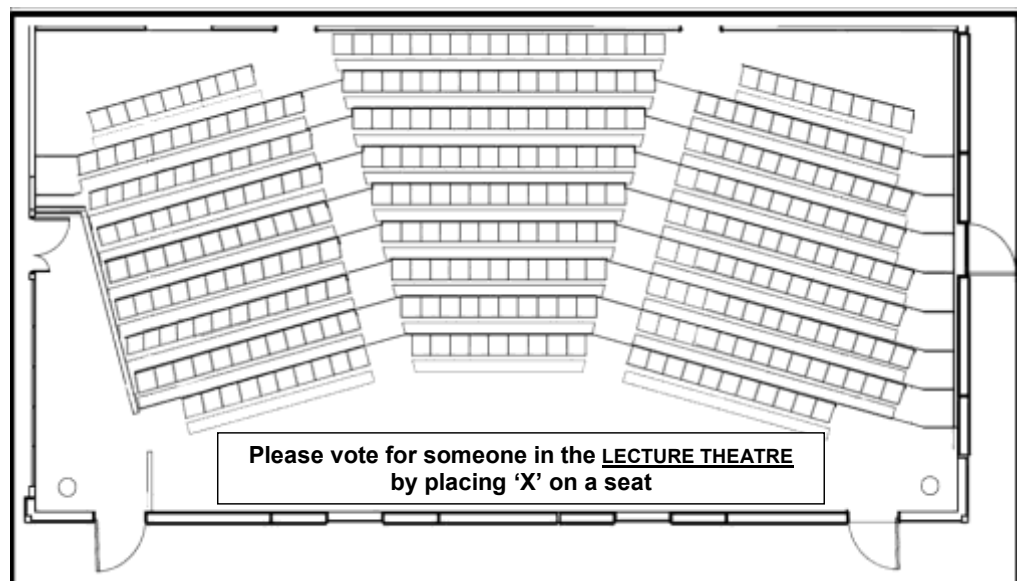
#### About YOU!

We would like to know a little bit about you so we can describe the demographics of our sample (please leave blank if you do not wish to disclose your gender/age). If you do not want to consent to take part in the study then simply do not make any votes. Even if you don't make any votes though, you will still gain a 'Credit Point' and voucher for the raffle so please put in your seat number. When you have finished please place this sheet back in the envelope provided. The whole process should only take a couple of minutes. There are no 'correct' answers and the only real guidance is that you can not vote for yourself.

I am female/male and I am \_\_\_\_\_ years old.

I am in Seat number:-

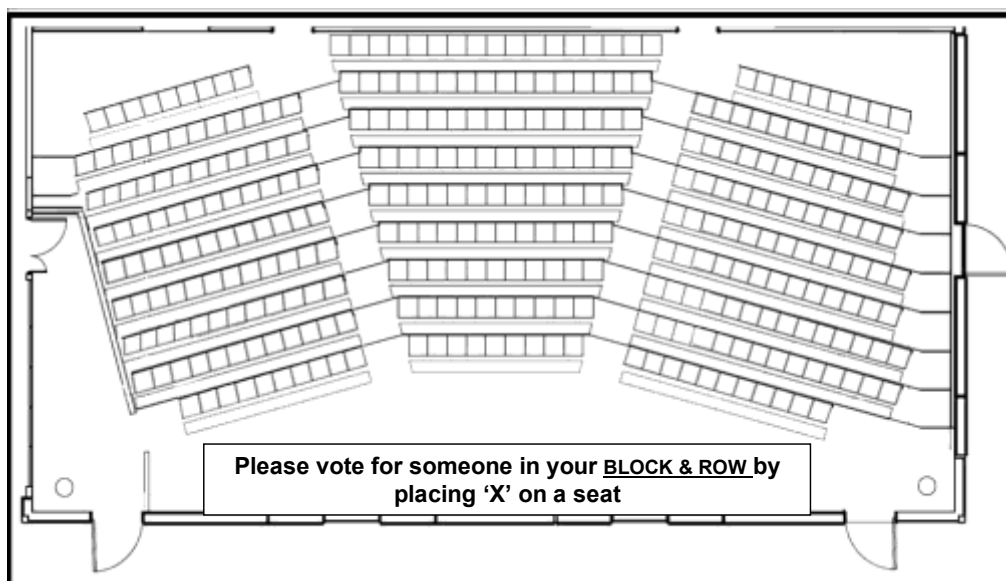
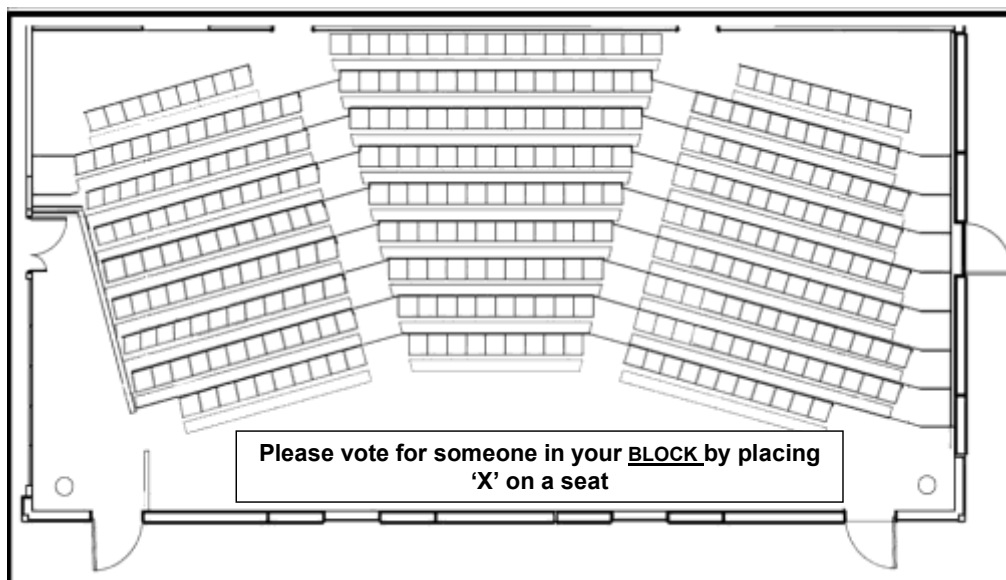
**IMPORTANT: Keep a personal record of your seat number!!**



**PTO**

## Appendix A10

Your vote is NEUTRAL and gives ZERO voucher



This final question is optional and open-ended about how you made your choice. I made my choices by:-

.....  
 .....  
 .....  
 .....  
 .....(optional)

THANK YOU! Paul, Patrick, Adrian, Zamira, Chris.

## Appendix B

### SHEET 1 Ethics, Background and Instructions

#### BACKGROUND AND INSTRUCTIONS

1. **Ethics - 'Credit points' and 'Lottery'**:- Any research that gets carried out in the University has to be approved by the relevant ethics committee. I have to ensure that everybody here:-
- i.) Is fully informed as to what the procedures are in advance.
  - ii.) Has the right to confidentiality/anonymity.
  - iii.) Gives their informed consent to take part.
  - iv.) Has the right to withdraw, without losing rewards/benefits.

Everybody here is entitled to **one** 'Credit Point'\* and everybody here will be entitled to a minimum of **one** ticket to a lottery\*. (Some people might get more than one ticket but nobody will get less than one.) The lottery prize will be.....

2. **Background**:- We have carried out a bit of research already that has looked at 'voting' behaviour where:-

- I.) 'voters' have had '*open*' information about whom to vote for and why,
- II.) and, the vote was '*open*' (the 'voter' announced their 'vote' publicly),
- III.) and, the vote was *POSITIVE* or *NEGATIVE*.

3. **Aim for Today**:- We want to collect some Control/Normative data today to use as a baseline for what we've done before, described in 1.) above, but

- I.) 'voters' have '*closed*' no information about whom to vote for and why, and
- II.) the 'vote' is *closed* (the 'voter' makes their vote privately).
- III.) and the vote will be *POSITIVE*, *NEGATIVE*\*\*.

4. **Procedure**:- We will ask you to make **THREE** votes that vary according to **spatial scale**.
- I.) Vote for someone in the **BLOCK** and in your **ROW** by making a cross on the seating plan.
  - II.) Vote for someone in your **BLOCK** by making a cross.
  - III.) Vote for someone in your **LECTURE THEATRE** by making a cross.
  - VI.) MOST IMPORTANTLY, please indicate your seat number here and make a note of it:-

I AM SAT IN BLOCK:-

I AM SAT IN ROW:-

I AM SAT IN SEAT:-

#### RULES

**Please be quiet until everyone has made their votes and the experimenter**  
**The only rule is that you can't vote for yourself!!!!**

5. **Thank you!!!!** Please make your judgements on the attached sheets and when you have finished just place your answer sheets in the envelope in which they came.

### About YOU!

M/F

Age

## Appendix B1

### Your vote is **NEGATIVE** and removes **FIVE** vouchers

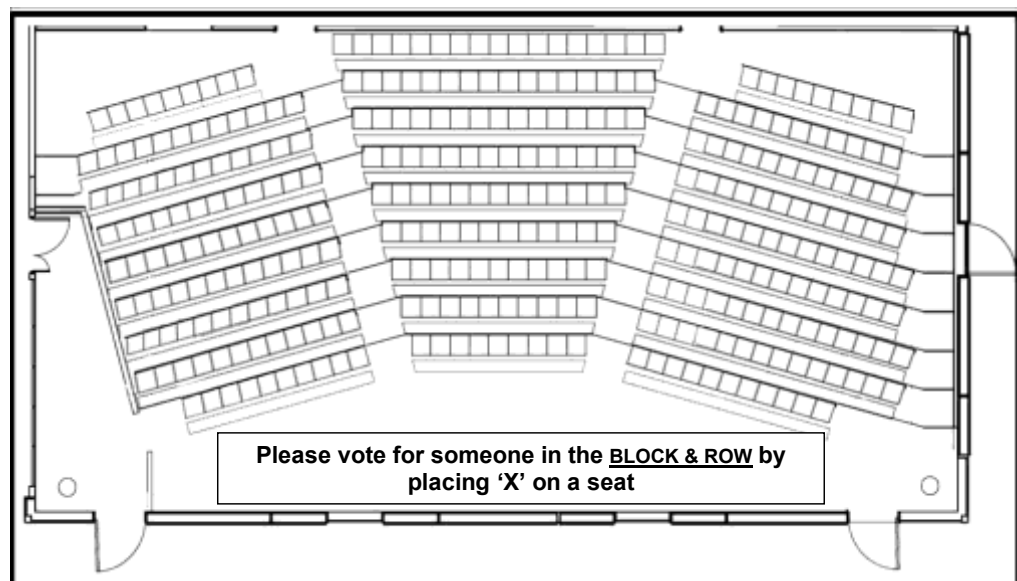
#### **About YOU!**

We would like to know a little bit about you so we can describe the demographics of our sample (please leave blank if you do not wish to disclose your gender/age). If you do not want to consent to take part in the study then simply do not make any votes. Even if you don't make any votes though, you will still gain a 'Credit Point' and voucher for the raffle so please put in your seat number. When you have finished please place this sheet back in the envelope provided. The whole process should only take a couple of minutes. There are no 'correct' answers and the only real guidance is that you can not vote for yourself.

I am female/male and I am \_\_\_\_\_ years old.

I am in Seat number:-

**IMPORTANT: Keep a personal record of your seat number!!**

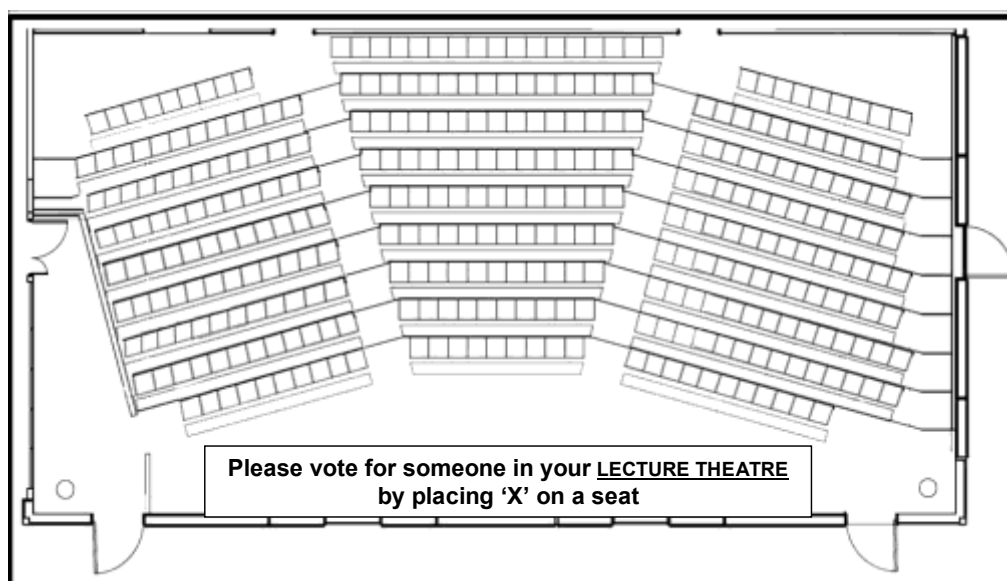
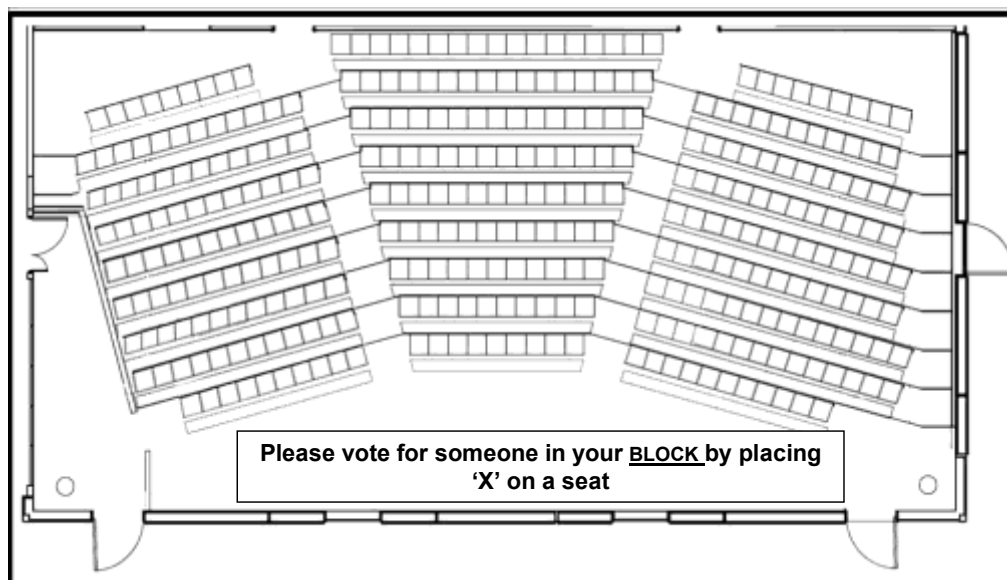


**PTO**



## Appendix B2

Your vote is NEGATIVE and removes FIVE vouchers



This final question is optional and open-ended about how you made your choice. I made my choices by:-

.....  
 .....  
 .....  
 .....  
 .....(optional)

THANK YOU! Paul, Patrick, Adrian, Zamira, Chris.

## Appendix B3

### Your vote is POSITIVE and adds FIVE vouchers

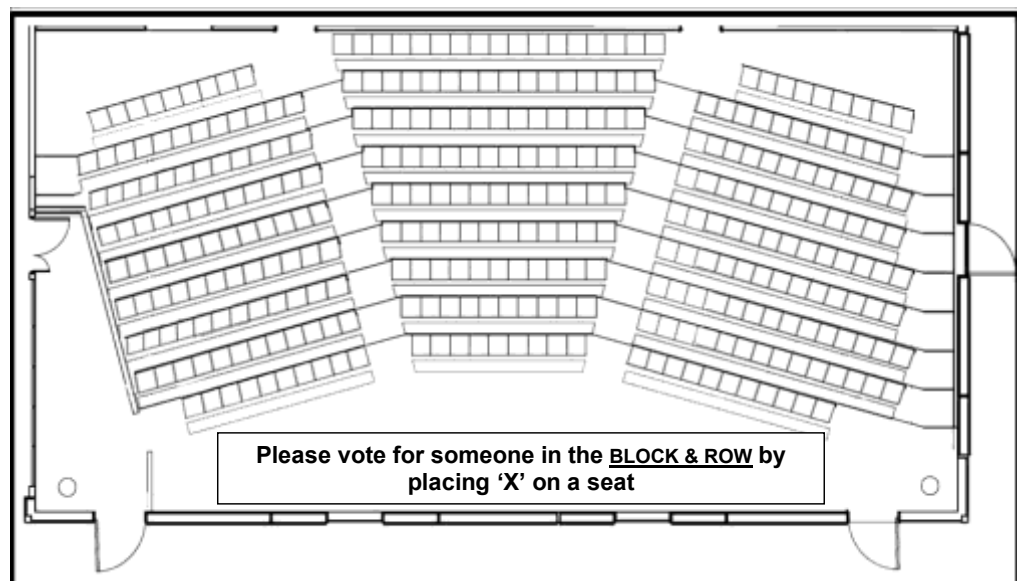
#### About YOU!

We would like to know a little bit about you so we can describe the demographics of our sample (please leave blank if you do not wish to disclose your gender/age). If you do not want to consent to take part in the study then simply do not make any votes. Even if you don't make any votes though, you will still gain a 'Credit Point' and voucher for the raffle so please put in your seat number. When you have finished please place this sheet back in the envelope provided. The whole process should only take a couple of minutes. There are no 'correct' answers and the only real guidance is that you can not vote for yourself.

I am female/male and I am \_\_\_\_\_ years old.

I am in Seat number:-

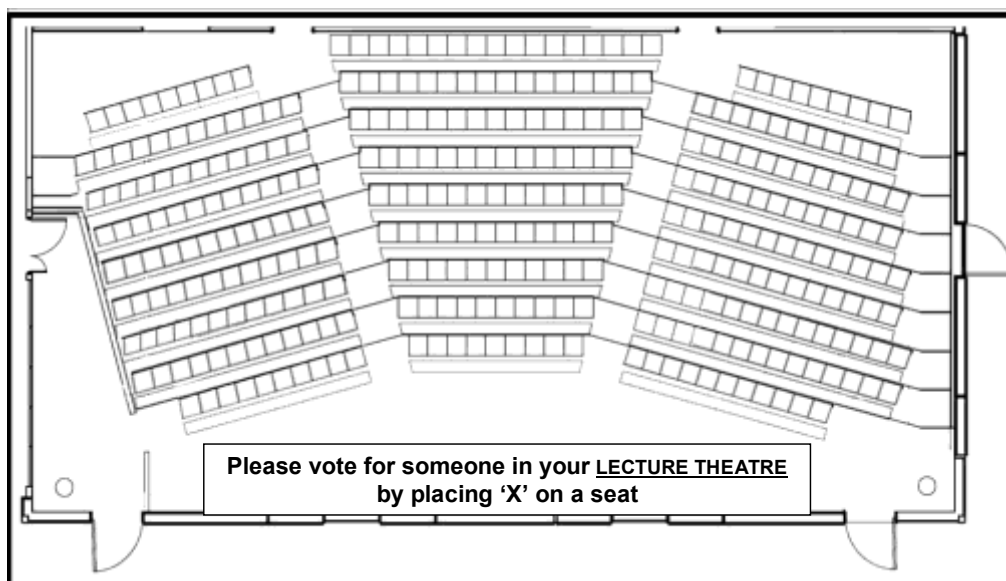
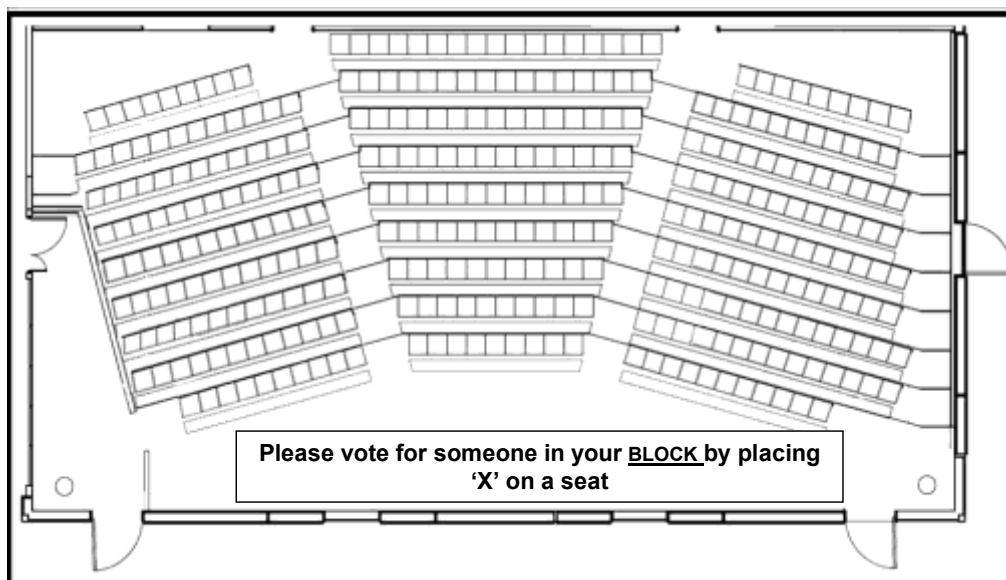
**IMPORTANT: Keep a personal record of your seat number!!**



**PTO**

## Appendix B4

Your vote is POSITIVE and adds FIVE vouchers



This final question is optional and open-ended about how you made your choice. I made my choices by:-

.....  
 .....  
 .....  
 .....  
 .....(optional)

THANK YOU! Paul, Patrick, Adrian, Zamira, Chris.

## Appendix C

### RESPONSE SHEET

#### **A little bit About YOU!**

We would like to know a little bit about you so we can describe the demographics of our sample (please leave blank if you do not wish to disclose your gender/age).

If you consent to taking part in the study then simply fill in this response sheet. If you don't want to take part then leave it blank, but please still put in your seat number, below, because you will still be entered to win a prize and still get a 'credit point'.

When you have finished please place this sheet back in the envelope provided. The whole process should only take a couple of minutes.

I am female/male and I am \_\_\_\_\_ years old.

I am in Seat number:-

**IMPORTANT: Keep a personal record of your seat number!!**

-----

**We are going to ask you to play the 'coin game' 3 times.**

**Each time you play, you will indicate how you play by circling 'H' or 'T' or entering it into a text box, on the other side of this 'Response Sheet'. Your responses are secret and anonymous and other players will not see what you have done. You will never get to see how other players responded.**

**Your aim is to try and score as many points as possible. We will work this out by adding up how many points you scored in the 3 games and dividing it by how many players there were that you played against. This will give us an average point score for each participant.**

**When you have made your responses we will give you a verbal debrief about what the study is about, but in short there are no catches or tricks involved. Computer scientists have generated computer programs that have played these kinds of games before. We want to know if a group of people play the game the same way as the computer program.**

**PTO**















## References

- Abeckaser, A. A., Golombek, F., Melita, P., Robbins, I., Schoof, A., & Singer, U. (Producers), & Almereyda, M. (Director). (2015). *Compliance*. United States: Magnolia Pictures.
- Antonovics, K., Arcidiacono, P., & Walsh, R. (2005). Games and discrimination lessons from the weakest link. *Journal of Human Resources*, 40(4), 918-947.
- Arendt, H. (1964). *Eichmann in Jerusalem* Viking Press New York.
- Ball, P. (2004) *Critical Mass: How One Thing Leads to Another*. London: William Heinemann.
- Banaji, M. R., & Greenwald, A. G. (1994). Implicit stereotyping and prejudice. *The Psychology of Prejudice: The Ontario Symposium*, 7 55-76.
- Barmish, R. B., & Boston, N. (2009). Risk and return considerations in “The weakest link”. *American Mathematical Monthly*, 116(4), 305-315.
- Barmish, R. B., & Boston, N. (2009). Risk and return considerations in “The weakest link”. *American Mathematical Monthly*, 116(4), 305-315.
- Baron, J. (1995). Blind justice: Fairness to groups and the do-no-harm principle. *Journal of Behavioural Decision Making*, 8(2), 71-83.

- Baumrind, D. (1964). Some thoughts on ethics of research: After reading milgram's" behavioural study of obedience.". *American Psychologist*, 19(6), 421 – 423.  
doi:10.1037/h0040128.
- Beauvois, J. L., Courbet, D., & Oberlé, D. (2012). The prescriptive power of the television host. A transposition of Milgram's obedience paradigm to the context of TV game-show. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 62(3), 111-119.
- Becker, Gary S. 1957. *The Economics of Discrimination*. Chicago, IL: University of Chicago Press.
- Bell, D. (1991). Modes of exchange: Gift and commodity. *The Journal of Socio-Economics*, 20(2), 155-167.
- Bellak, G (Producer), & Dubin, C. S. (Director). (1976). *The Tenth Level*. United States: Paramount Pictures.
- Belot, M., Bhaskar, V., & van de Ven, J. (2010). Promises and cooperation: Evidence from a TV game-show. *Journal of Economic Behaviour & Organization*, 73(3), 396-405.
- Blass, T. (1991). Understanding behaviour in the milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology*, 60(3), 398.

- Blass, T. (1996). Attribution of responsibility and trust in the milgram obedience experiment. *Journal of Applied Social Psychology*, 26, 1529-1535.
- Blass, T. (1999). The milgram paradigm after 35 years: Some things we now know about obedience to authority<sup>1</sup>. *Journal of Applied Social Psychology*, 29(5), 955-978.
- Blass, T. (2002). THE MAN WHO SHOCKED THE WORLD controversial social psychologist stanley milgram uncovered a profound and disturbing truth about human nature. *Psychology Today-New York-*, 35(2), 68-74.
- Blass, T. (2004). *The man who shocked the world: The life and legacy of stanley milgram* Basic Books New York.
- Blass, T. (2009). From new haven to santa clara: A historical perspective on the milgram obedience experiments. *American Psychologist*, 64(1), 37.
- Blass, T., & Schmitt, C. (2001). The nature of perceived authority in the milgram paradigm: Two replications. *Current Psychology*, 20(2), 115-121.
- Brannigan, A. (2013). *Beyond the banality of evil: Criminology and genocide* Oxford University Press.
- Bressen, T. (2007). Consensus decision making. *The change handbook: The definitive resource on today's best methods for engaging whole systems*, 212-217.

- Brewer, M. B. (2000). Research design and issues of validity. *Handbook of research methods in social and personality psychology*, 3-16.
- Brown, R. (1986). *Social psychology roger brown* (2nd ed ed.) New York Free Press  
Lonon Collier Macmillan.
- Bryman, A. (1984). The debate about quantitative and qualitative research: A question of method or epistemology? *British Journal of Sociology*, , 75-92.
- Burger, J. M. (2009). Replicating milgram: Would people still obey today? *American Psychologist*, 64(1), 1.
- Burger, J. M., Girgis, Z. M., & Manning, C. C. (2011). In their own words: Explaining obedience to authority through an examination of participants' comments. *Social Psychological and Personality Science*, 1948550610397632.
- Cadsby, B., Du, N., & Song, F. (2016). In-group favoritism and moral decision-making. *Journal of Economic Behaviour & Organization*.
- Coe, Robert. (2002, September). *It's the effect size, stupid, what effect size is and why it is important*. Paper presented at the Annual Conference of the British Educational Research Association, Exeter, United Kingdom. Retrieved from <https://www.leeds.ac.uk/educol/documents/00002182.htm>

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.

doi:10.1037/0033-2909.112.1.155

Correll, J., Judd, C. M., Park, B., Wittenbrink, B., Dovidio, J., Hewstone, M., . . . Esses, V.

(2010). Measuring prejudice, stereotypes and discrimination. *The Sage Handbook of Prejudice, Stereotyping and Discrimination*. London: SAGE, , 45-62.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56(1), 5.

Dickinson, R. (2002). The Milgram re-enactment [Performance]. London, UK: South London Gallery.

Dolan, K. (1998). Voting for Women in the "Year of the Woman". *American Journal of Political Science*, 272-293.

Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. *Advances in Experimental Social Psychology*, 36, 1-52.

Downs, G. W., Rocke, D. M., & Barsoom, P. N. (1996). Is the good news about compliance good news about cooperation? *International Organization*, 50(03), 379-406.

Doyle, A.C. (1890). *The sign of the four*. Penguin Classic.

- Eyres, A., Bowie, B., Smithson, J., Winslow, S., & Day, P.J. (Producers) & Driscoll, Z. & Murphy, A. A. (Director). (2011, October 30). *Curiosity: How evil are you?* [Television broadcast]. California: Darlow Smithson Production & Edge West Production.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137-140.
- Fehr, E., Tougareva, E., & Fischbacher, U. (2014). Do high stakes and competition undermine fair behaviour? evidence from russia. *Journal of Economic Behaviour & Organization*, 108, 354-363.
- Février, P., & Linnemer, L. (2006). Equilibrium selection: Payoff or risk dominance?: The case of the “Weakest link”. *Journal of Economic Behaviour & Organization*, 60(2), 164-181.
- Fletcher, M., & Mooney, C. (2003). The weakest link. *Teaching Statistics*, 25(2), 54-55.  
doi:10.1111/1467-9639.00125
- Fong, C. (2001). Social preferences, self-interest, and the demand for redistribution. *Journal of Public Economics*, 82, 225-246. doi:10.1016/S0047-2727(00)00141-9



- Furnham, A., Kidwai, A., & Thomas, C. (2001). Personality, psychometric intelligence, and self-estimated intelligence. *Journal of Social Behaviour and Personality*, 16(1), 97-114.
- Gaertner, S. L., & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly*, 23-30.
- Gardin, H., Kaplan, K. J., Firestone, I. J., & Cowan, G. A. (1973). Proxemic effects on cooperation, attitude, and approach-avoidance in a prisoner's dilemma game. *Journal of Personality and Social Psychology*, 27(1), 13.
- Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *QUARTERLY JOURNAL OF ECONOMICS-CAMBRIDGE MASSACHUSETTS*, 118(3), 1049-1074.
- Goddard, P., Ashley, R., Fuller, G., & Hudson, I. (2011, July). Love thy neighbour: proxemic bias in the voting strategy of contestants in the TV quiz-show "The Weakest Link". In *Proceedings of the 2011 Conference of the International Confederation for the Advancement of Behavioural Economics and Economic Psychology*. Washington Singer Press.
- Goddard, Paul (2012, July) *Bias for proximity and gender in the voting patterns of contestants in the TV quiz-show 'The Weakest Link'*. In: SABE 2012, 12 - 15 July 2012, Granada, Spain.

- Goddard, P., Hylton, P., Parke, A. and Noh, Z. (2013, July) *Tit-for-Tat voting by contestants in the TV quiz-show 'The Weakest Link'*. In: SABE/IAREP/ICABEEP 2013 Conference in Atlanta, GA., 25-29 July 2013, Atlanta, Georgia.
- Goffman, E. (1983). The interaction order: American sociological association, 1982 presidential address. *American Sociological Review*, 48(1), 1-17.
- Goldberg, P. (1968). Are women prejudiced against women? *Society*, 5(5), 28-30.
- Gordon, D. S., Madden, J. R., & Lea, S. E. (2014). Both loved and feared: third party punishers are viewed as formidable and likeable, but these reputational benefits may only be open to dominant individuals. *PloS one*, 9(10), e110045.
- Gordon, D. S., & Lea, S. E. (2016). Who Punishes? The Status of the Punishers Affects the Perceived Success of, and Indirect Benefits From, "Moralistic" Punishment. *Evolutionary Psychology*, 14(3), 1474704916658042.
- Gottman, J. M. (2011). *The science of trust: Emotional attunement for couples* WW Norton & Company.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009).

Understanding and using the implicit association test: III. meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17.

Haan, M. A., Los, B., & Riyanto, Y. E. (2011). Signaling strength? An analysis of decision making in The Weakest Link. *Theory and decision*, 71(4), 519-537.

Hall, E. T. (1963). A system for the notation of proxemic behaviour. *American Anthropologist*, 65(5), 1003-1026.

Hall, E. T., Birdwhistell, R. L., Bock, B., Bohannon, P., Diebold Jr, A. R., Durbin, M., . . . Kimball, S. T. (1968). Proxemics [and comments and replies]. *Current Anthropology*, , 83-108.

Hartley, R., Lanot, G., & Walker, I. (2014). Who really wants to be a millionaire? Estimates of risk aversion from gameshow data. *Journal of Applied Econometrics*, 29(6), 861-879.

Heath, A. F., & McMahon, D. (2005). 14 social mobility of ethnic minorities. *Ethnicity, Social Mobility, and Public Policy: Comparing the USA and UK*, , 393.

Herrmann, B., Thoni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science (New York, N.Y.)*, 319(5868), 1362-1367.  
doi:10.1126/science.1153808 [doi]

- Hertwig, R. (2012). The psychology and rationality of decisions from experience. *Synthese*, 187(1), 269-292.
- De Heus, P., Hoogervorst, N., & Van Dijk, E. (2010). Framing prisoners and chickens: Valence effects in the prisoner's dilemma and the chicken game. *Journal of Experimental Social Psychology*, 46(5), 736-742.
- Home Office. (February, 1999). *The Stephen Lawrence Inquiry: Report of an Inquiry by Sir William Macpherson of Cluny*, Cm 4262-I.
- Iachini, T., Pagliaro, S., & Ruggiero, G. (2015). Near or far? it depends on my impression: Moral information and spatial behaviour in virtual interactions. *Acta Psychologica*, 161, 131-136. doi:10.1016/j.actpsy.2015.09.003
- Inderbitzin, M. P., Betella, A., Lanatá, A., Scilingo, E. P., Bernardet, U., & Verschure, P. F. M. J. (2013). The social perceptual salience effect. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 62-74. doi:10.1037/a0028317; 10.1037/a0028317.supp (Supplemental)
- Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, 25(6), 881-919.
- Kahneman, D. (2012). *Thinking, fast and slow* London Penguin 2012.

- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive psychology*, 3(3), 430-454.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237.
- Kahneman, D., & Tversky, A. (1979). On the interpretation of intuitive probability: A reply to jonathan cohen. *Cognition*, 7(4), 409-411.
- Kilham, W., & Mann, L. (1974). Level of destructive obedience as a function of transmitter and executant roles in the milgram obedience paradigm. *Journal of Personality and Social Psychology*, 29(5), 696.
- Kocher, M. G., Martinsson, P., & Visser, M. (2008). Does stake size matter for cooperation and punishment? *Economics Letters*, 99(3), 508-511.
- Leliveld, M. C., Van Beest, I., Van Dijk, E., & Tenbrunsel, A. E. (2009). Understanding the influence of outcome valence in bargaining: A study on fairness accessibility, norms, and behaviour. *Journal of Experimental Social Psychology*, 45(3), 505-514.
- Levitt, S. D. (2004). Testing theories of discrimination: evidence from Weakest Link. *The Journal of Law and Economics*, 47(2), 431-452.

- Matthews, J. L., & Matlock, T. (2011). Understanding the link between spatial distance and social distance. *Social Psychology*, 42 (3), 185 – 192.
- McCall, C., & Singer, T. (2015). Facing off with unfair others: Introducing proxemic imaging as an implicit measure of approach and avoidance during social interaction. *PloS One*, 10(2), e0117532.
- Meeus, W. H., & Raaijmakers, Q. A. (1995). Obedience in modern society: The utrecht studies. *Journal of Social Issues*, 51(3), 155-175.
- Metrick, A. (1995). A natural experiment in "jeopardy!". *The American Economic Review*, , 240-253.
- Milgram, S. (1963). Behavioural study of obedience. *The Journal of Abnormal and Social Psychology*, 67(4), 371-378. doi:10.1037/h0040525
- Milgram, S. (1964). Issues in the study of obedience: A reply to baumrind. *American Psychologist*, 19 (11), 848-852. doi:10.1037/h0044954
- Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations*, 18(1), 57-76.
- Mills, S. (Producer) & Caron. B. (Director). (2006, January 4). *The Heist* [Television broadcast]. London: Channel 4.

- Nieuwenhuis, J., Völker, B., & Flap, H. (2013). "A bad neighbour is as great a plague as a good one is a great blessing": On negative relationships between neighbours. *Urban Studies*, 0042098013482508.
- Newcomb, T. M. (1956). The prediction of interpersonal attraction. *American Psychologist*, 11(11), 575.
- Noh, Z., Goddard, P., Hylton, P., & Parke, A. (2014, September). Voting bias: Switches in the neighbour effect as a function of vote valence. *International Congress of Applied Psychology (ICAP) 2014*.
- Osborne, M. J. (2009). *An introduction to game theory* (International ed.). New York: Oxford University Press.
- Parke, J., & Parke, A. (2013). Does size really matter? A review of the role of stake and prize levels in relation to gambling-related harm. *The Journal of Gambling Business and Economics*, 7(3), 77-110.
- Pasquaretta, C., Battesti, M., Klenschi, E., Bousquet, C. A., Sueur, C., & Mery, F. (2016, March). How social network structure affects decision-making in *Drosophila melanogaster*. In *Proc. R. Soc. B* (Vol. 283, No. 1826, p. 20152954). The Royal Society.

- Perry, G. (2012). Beyond the shock machine: The untold story of the milgram obedience experiments. *Melbourne, Vic.: Scribe*,
- Post, T., Van den Assem, M. J., Baltussen, G., & Thaler, R. H. (2008). Deal or no deal? decision making under risk in a large-payoff game-show. *The American economic review*, 38-71.
- Poundstone, W. (1992). Prisoner's dilemma: John von neuman, game theory, and the puzzle of the bomb.
- Powell. (1984). *The kindly ones* Fontana Paperbacks.
- Pras, B., & Summers, J. O. (1978). Perceived risk and composition models for multiattribute decisions. *Journal of Marketing Research*, , 429-437.
- Quillian, L. (2006). New approaches to understanding racial prejudice and discrimination. *Annual Review of Sociology*, , 299-328.
- Rai, H., & Kulkarni, V. (2008). How close can you get: The effect of proxemics on leader-member exchange quality. *ICFAI Journal of Management Research*, 7(9), 7-15.
- Sally, D. (2001). On sympathy and games. *Journal of Economic Behaviour & Organization*, 44(1), 1-30.



- Scheinkman, J. A. (2008). Social interactions. *The New Palgrave Dictionary of Economics*, 2.
- Schlosser, J. (2001). Weakest link no ratings weakling. *Broadcasting & Cable*, 131(17), 14.
- Sharifpour, M., & Walters, G. (2014). The interplay between prior knowledge, perceived risk and the tourism consumer decision process: A conceptual framework. *The Marketing Review*, 14(3), 279-296.
- Shaw, J. I., Bergen, J. E., Brown, C. A., & Gallagher, M. E. (2000). Centrality preferences in choices among similar options. *The Journal of general psychology*, 127(2), 157-164.
- Sheridan, C. L., & King, R. G. (1972). Obedience to authority with an authentic victim. *Proceedings of the Annual Convention of the American Psychological Association*,
- Simcock, P., Sudbury, L., & Wright, G. (2006). Age, perceived risk and satisfaction in consumer decision making: A review and extension. *Journal of Marketing Management*, 22(3-4), 355-377.
- Simon, H. A. (1955). A behavioural model of rational choice. *The Quarterly Journal of Economics*, , 99-118.

- Sip, K. E., Skewes, J. C., Marchant, J. L., McGregor, W. B., Roepstorff, A., & Frith, C. D. (2015). What if I get busted? Deception, choice, and decision-making in social interaction. *Decision-Making Experiments under a Philosophical Analysis: Human Choice as a Challenge for Neuroscience*, 8.
- Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., . . . Sanchez-Vives, M. V. (2006). A virtual reprise of the stanley milgram obedience experiments. *PloS One*, 1(1), e39.
- Stangneth, B. (2014). *Eichmann before jerusalem: The unexamined life of a mass murderer* Random House.
- Stanley, M. (1974). Obedience to authority. *An Experimental View*. Harper, New York,
- Strandburg-Peshkin, A., Farine, D. R., Couzin, I. D., & Crofoot, M. C. (2015). Shared decision-making drives collective movement in wild baboons. *Science*, 348(6241), 1358-1361.
- Sullivan, G. M., & Feinn, R. (2012). Using Effect Size—or Why the *P* Value Is Not Enough. *Journal of Graduate Medical Education*, 4(3), 279–282.  
<http://doi.org/10.4300/JGME-D-12-00156.1>

- Tamir, M., Bigman, Y. E., Rhodes, E., Salerno, J., & Schreier, J. (2015). An expectancy-value model of emotion regulation: Implications for motivation, emotional experience, and decision making. *Emotion, 15*(1), 90.
- Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*(2), 440.
- Tucker, A. W. (1983). The mathematics of tucker: A sampler. *The Two-Year College Mathematics Journal, 14*(3), 228-232.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207-232.
- Tversky, A., & Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases. *Utility, probability, and human decision making* (pp. 141-162) Springer.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297-323.
- Verplanken, B., & Holland, R. W. (2002). Motivated decision making: Effects of activation and self-centrality of values on choices and behaviour. *Journal of Personality and Social Psychology, 82*(3), 434-447. doi:10.1037/0022-3514.82.3.434

- Van Dolder, D., Van den Assem, M. J., Camerer, C., & Thaler, R. H. (2015, May). Standing united or falling divided? High stakes bargaining in a TV game-show. In *American Economic Review, Papers and Proceedings* (Vol. 105, No. 5, pp. 402-407).
- Weber, M., & Camerer, C. F. (1998). The disposition effect in securities trading: An experimental analysis. *Journal of Economic Behaviour & Organization*, 33(2), 167-184. doi:[http://dx.doi.org/10.1016/S0167-2681\(97\)00089-9](http://dx.doi.org/10.1016/S0167-2681(97)00089-9)
- Whitney, S. N. (2003). A new model of medical decisions: exploring the limits of shared decision making. *Medical Decision Making*, 23(4), 275-280.
- Valenzuela, A., & Raghurir, P. (2007). The role of strategy in mixed-gender group interactions: A study of the television show "The weakest link". *Sex Roles*, 57(3-4), 293-303.
- Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology*, 10(2), 109-120.
- Zobel, C., Lin, S., Sena, T., Muskat, L., & Davidson, T. (Producer), & Zobel, C. (Director). (2012). *Compliance*. United States: Magnolia Pictures.